



## TESIS

# PEMILIHAN TITIK KNOT DAN BANDWIDTH YANG OPTIMAL MENGGUNAKAN CROSS-VALIDATION (CV), GENERALIZED CROSS-VALIDATION (GCV) DAN UNBIASED RISK (UBR) PADA PEMODELAN REGRESI NONPARAMETRIK ESTIMATOR CAMPURAN SPLINE TRUNCATED DAN KERNEL GAUSSIAN

**ANDREA TRI RIAN DANI**

**06211950015001**

**Dosen Pembimbing:**

**Dr. Vita Ratnasari, M.Si**

**Prof. Dr. Drs. I Nyoman Budiantara, M.Si**

**Departemen Statistika  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember  
2021**



**TESIS**

**PEMILIHAN TITIK KNOT DAN BANDWIDTH YANG OPTIMAL MENGGUNAKAN CROSS-VALIDATION (CV), GENERALIZED CROSS-VALIDATION (GCV) DAN UNBIASED RISK (UBR) PADA PEMODELAN REGRESI NONPARAMETRIK ESTIMATOR CAMPURAN SPLINE TRUNCATED DAN KERNEL GAUSSIAN**

**ANDREA TRI RIAN DANI**

**06211950015001**

**Dosen Pembimbing:**

**Dr. Vita Ratnasari, M.Si**

**Prof. Dr. Drs. I Nyoman Budiantara, M.Si**

**Departemen Statistika**

**Fakultas Sains dan Analitika Data**

**Institut Teknologi Sepuluh Nopember**

**2021**



## **THESIS**

# **OPTIMAL SELECTION OF KNOT POINTS AND BANDWIDTH USING CROSS-VALIDATION (CV), GENERALIZED CROSS-VALIDATION (GCV) AND UNBIASED RISK (UBR) IN MODELING MIXED ESTIMATOR NONPARAMETRIC REGRESSION TRUNCATED SPLINE AND GAUSSIAN KERNEL**

**ANDREA TRI RIAN DANI**

**06211950015001**

**Supervisors:**

**Dr. Vita Ratnasari, M.Si**

**Prof. Dr. Drs. I Nyoman Budiantara, M.Si**

**Department of Statistics**

**Faculty of Science and Data Analytics**

**Institut Teknologi Sepuluh Nopember**

**2021**



## LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar

**Magister Statistika (M.Stat)**

di

**Institut Teknologi Sepuluh Nopember**

Oleh:

**ANDREA TRI RIAN DANI**  
**NRP. 06211950015001**

Tanggal Ujian : 01 Maret 2021  
Periode Wisuda : April 2021

Disetujui Oleh:

**Tim Pembimbing**

1. Dr. Vita Ratnasari, M.Si  
NIP: 19700910 199702 2 001

*Ratnasari*

2. Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
NIP: 19650603 198903 1 003

*Budiantara*

**Tim Penguji**

1. Dr. rer.pol. Dedy Dwi Prastyo, M.Si  
NIP: 19831204 200812 1 002

*Prastyo*

2. Santi Wulan Purnami, M.Si, P.hD  
NIP: 19720923 199803 2 001

*Purnami*



Kepala Departemen Statistika  
Fakultas Sains dan Analitika Data

Dr. Dra. Kartika Fithriasari, M.Si.  
NIP: 19691212 199303 2 002

*Fithriasari*



*(Halaman ini sengaja dikosongkan)*

**PEMILIHAN TITIK KNOT DAN BANDWIDTH YANG OPTIMAL  
MENGUNAKAN CROSS-VALIDATION (CV), GENERALIZED CROSS-  
VALIDATION (GCV) DAN UNBIASED RISK (UBR) PADA PEMODELAN  
REGRESI NONPARAMETRIK ESTIMATOR CAMPURAN  
SPLINE TRUNCATED DAN KERNEL GAUSSIAN**

Nama Mahasiswa : Andrea Tri Rian Dani  
NRP : 06211950015001  
Pembimbing : Dr. Vita Ratnasari, M.Si  
Co-Pembimbing : Prof. Dr. Drs. I Nyoman Budiantara, M.Si

**ABSTRAK**

Pada regresi nonparametrik *spline truncated* menentukan titik knot optimal menjadi hal yang sangat penting dan krusial. Hal yang sama berlaku pada regresi kernel, dimana hal yang harus diperhatikan adalah pemilihan *bandwidth*. Maka menjadi permasalahan tersendiri apabila menggunakan model estimator campuran, khususnya mengkombinasikan estimator *spline truncated* dengan kernel Gaussian. Penentuan titik knot dan *bandwidth* dalam model regresi nonparametrik estimator campuran akan sangat mempengaruhi kurva regresi yang akan dibentuk. Terdapat beberapa metode yang dapat digunakan untuk memilih titik knot dan *bandwidth* yang optimal, yaitu *Cross-Validation* (CV), *Generalized Cross-Validation* (GCV) dan *Unbiased Risk* (UBR). Tujuan penelitian ini adalah mendapatkan bentuk estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian, kemudian membandingkan kinerja pemilihan titik knot dan *bandwidth* menggunakan metode CV, GCV dan UBR pada data simulasi. Tujuan ketiga adalah membandingkan hasil pemilihan titik knot dan *bandwidth* yang optimal menggunakan metode CV, GCV dan UBR pada data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019. Metode estimasi yang digunakan adalah *Ordinary Least Squares* (OLS). Kajian simulasi dilakukan dengan membangkitkan fungsi trigonometri untuk komponen *spline truncated* dan kernel Gaussian dengan *error* yang mengikuti distribusi Normal, serta pada kombinasi ukuran sampel dan varians.

Berdasarkan hasil analisis, kajian simulasi menunjukkan metode GCV memberikan performa hasil serta ketepatan yang lebih baik untuk setiap kombinasi ukuran sampel dan variasi varians dalam menduga titik knot dan *bandwidth* optimal. Hasil aplikasi pada data persentase penduduk miskin menunjukkan hal yang sama, dimana metode GCV lebih baik dalam menduga titik knot dan *bandwidth* optimal dibandingkan metode CV maupun UBR. Metode CV menggunakan tiga titik knot dan *bandwidth* optimal didapatkan nilai  $R^2$  sebesar 83,80%. Pada metode UBR menggunakan tiga titik knot dan *bandwidth* optimal, didapatkan nilai  $R^2$  sebesar 79,24%. Metode GCV menggunakan tiga titik knot dan *bandwidth* optimal, didapatkan nilai  $R^2$  sebesar 90,52%.

**Kata Kunci** : *Cross-Validation* (CV), *Generalized Cross-Validation* (GCV), *Unbiased Risk* (UBR) dan Estimator Campuran

*(Halaman ini sengaja dikosongkan)*

**OPTIMAL SELECTION OF KNOT POINT AND BANDWIDTH  
USING CROSS-VALIDATION (CV), GENERALIZED CROSS-VALIDATION  
(GCV) AND UNBIASED RISK (UBR) IN MODELING  
MIXED ESTIMATOR NONPARAMETRIC REGRESSION  
TRUNCATED SPLINE AND GAUSSIAN KERNEL**

Student Name : Andrea Tri Rian Dani  
Student ID : 06211950015001  
Supervisor : Dr. Vita Ratnasari, M.Si  
Co-Supervisor : Prof. Dr. Drs. I Nyoman Budiantara, M.Si

**ABSTRACT**

*In nonparametric regression truncated spline, determining the optimal knot point is very important and crucial. The same thing applies to kernel regression, where the thing that must be considered is the choice of bandwidth. So it becomes a separate problem when using a mixed estimator model, in particular combining the truncated spline estimator with the Gaussian kernel. Determination of knots and bandwidth points in the mixed estimator nonparametric regression model will greatly affect the regression curve to be formed. There are several methods that can be used to select the optimal knot point and bandwidth, namely Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR). The purpose of this study is to obtain a mixed estimator form of nonparametric regression truncated spline and Gaussian kernel, then compare the performance of selecting knots and bandwidth points using CV, GCV, and UBR methods on the simulation data. The third objective is to compare the results of selecting the optimal knot point and bandwidth using the CV, GCV, and UBR methods on the data on the percentage of poor people in Districts/Cities in Borneo Island in 2019. The estimation method used is Ordinary Least Squares (OLS). Simulation studies are carried out by generating trigonometric functions for the truncated spline component and Gaussian kernel with an error following the Normal distribution, as well as on a combination of sample size and variance.*

*Based on the analysis results, the simulation study shows that the GCV method provides better result performance and accuracy for each sample size combination and variance variation in estimating optimal knot points and bandwidth. The application results on the percentage of poor people show the same thing, where the GCV method is better at estimating the optimal knot point and bandwidth than the CV and UBR methods. The CV method uses three-point knots and optimal bandwidth, and the  $R^2$  value is 83.80%. In the UBR method using three-point knots and optimal bandwidth,  $R^2$  value is 79.23%. The GCV method uses three-point knots and optimal bandwidth, the  $R^2$  value is 90.52%.*

**Keywords** : Cross-Validation (CV), Generalized Cross-Validation (GCV), Unbiased Risk (UBR) and Mixed Estimators



*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR



*Assalamu'alaikum Warrahmatulahi Wabarakatuh*

Alhamdulillahirobbil'alamin, segala puji dan syukur penulis panjatkan kehadirat Allah SWT, karena atas limpahan rahmat, hidayah serta pertolongan-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul “**Pemilihan Titik Knot dan Bandwidth Optimal menggunakan Cross-Validation (CV), Generalized Cross-Validation (GCV) dan Unbiased Risk (UBR) pada Pemodelan Regresi Nonparametrik Estimator Campuran Spline Truncated dan Kernel Gaussian**”.

Tesis ini ditulis untuk memenuhi syarat dalam menyelesaikan pendidikan Pascasarjana (S2) di Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Penulis menyadari selama penyusunan tesis ini tidak terlepas dari bantuan, bimbingan dan motivasi dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Vita Ratnasari, M.Si sebagai Sekretaris Departemen II (Bidang Sumber Daya Keuangan, Sumber Daya Manusia dan Sarana Prasarana) sekaligus sebagai Dosen Pembimbing 1 serta Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si sebagai Dosen Pembimbing 2 yang telah banyak meluangkan waktu untuk memberikan bimbingan, saran, arahan, motivasi dan ilmu yang bermanfaat dalam menyelesaikan tesis ini.
2. Bapak Dr. rer.pol. Dedy Dwi Prastyo, M.Si sebagai Kepala Program Studi Pascasarjana Departemen Statistika Fakultas Sains dan Analitika Data ITS sekaligus sebagai Dosen Penguji 1 dan Ibu Santi Wulan Purnami, M.Si, P.hD sebagai Sekretaris Departemen I (Bidang Akademik, Kemahasiswaan, Penelitian dan Pengabdian Kepada Masyarakat) dan juga sebagai Dosen Penguji 2 yang telah memberikan arahan, kritik, saran, serta masukan yang membangun demi kesempurnaan tesis ini.

3. Ibu Dr. Kartika Fithriasari, M.Si sebaga Kepala Departemen Statistika Fakultas Sains dan Analitika Data ITS.
4. Bapak Dr. Muhammad Mashuri, M.T sebagai dosen wali yang telah memberikan motivasi, serta masukan demi kelancaran dan terselesaikannya studi.
5. Bapak dan Ibu dosen pengajar di Departemen Statistika Fakultas Sains dan Analitika Data ITS, terima kasih atas ilmu yang telah diajarkan kepada penulis.
6. Bapak dan Ibu Staff Pegawai di Departemen Statistika FSAD ITS atas pelayanannya yang tulus dan ramah selama perkuliahan.
7. Terima kasih kepada kedua orang tua (H. Mariadi dan Hj. Sanatun), keluarga tercinta kakak kandung (Rian Bede T. D, S.Pd dan Andrika Dwi R. D, S.Pd), kakak ipar (Fendi I. S.Pd dan Eka S. S.Pd) dan keponakan tersayang (Xandra A. Irawan) serta seluruh keluarga yang selalu memberikan doa, motivasi dan semua yang terbaik kepada penulis.
8. Terima kasih kepada sahabat seperjuangan (Aliftha Ainurrochmah, Ludia Ni'matuzzahroh, Narita Y. A., Ardiana F. Dewi, Patrica Pungky. G., Ibu Riry S., Hartina H., Raditya N., I Gusti B. N. Diksa, Andi Rezky P. S., M. Alifian Nuriman) dan seluruh teman-teman seperjuangan S2 angkatan 2019 yang tidak dapat disebutkan satu persatu, terima kasih telah memberikan doa, dukungan, motivasi dan semangat kepada penulis.
9. Akhirnya terima kasih kepada setiap pihak yang tidak dapat disebutkan satu-persatu namanya, namun telah banyak membantu penulis dalam berbagai hal baik doa, motivasi dan lain sebagainya.

Semoga tesis ini dapat bermanfaat dan dapat menambah wawasan keilmuan. Penulis menyadari bahwa tesis ini masih jauh dari kata sempurna. Saran dan kritik yang membangun dari semua pihak, sangat penulis harapkan untuk perbaikan

Surabaya, 16 Januari 2021

Penulis

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN .....	iii
ABSTRAK .....	v
ABSTRACT .....	vii
KATA PENGANTAR .....	ix
DAFTAR ISI .....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR .....	xvii
DAFTAR LAMPIRAN .....	xix
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	7
1.3 Tujuan Penelitian .....	8
1.4 Manfaat Penelitian .....	8
1.5 Batasan Masalah .....	9
<b>BAB 2 TINJAUAN PUSTAKA.....</b>	<b>11</b>
2.1 Analisis Regresi .....	11
2.2 Regresi Parametrik dan Nonparametrik.....	12
2.3 Regresi Nonparametrik <i>Spline Truncated</i> .....	13
2.4 Regresi Nonparametrik Kernel .....	16
2.5 Estimator Campuran Regresi Nonparametrik.....	19
2.6 Pemilihan Titik Knot dan <i>Bandwidth</i> Optimal .....	21
2.6.1 Metode <i>Cross-Validation</i> (CV) .....	23
2.6.2 Metode <i>Generalized Cross-Validation</i> (GCV).....	23
2.6.3 Metode <i>Unbiased Risk</i> (UBR).....	24
2.7 Kriteria Pemilihan Model Terbaik.....	25
2.8 Kolmogorov-Smirnov Test.....	25
2.9 <i>Variance Inflation Factor</i> .....	26

2.10	Tinjauan Kemiskinan .....	26
2.10.1	Definisi Kemiskinan .....	27
2.10.2	Konsep dan Pengukuran Kemiskinan .....	28
2.10.3	Faktor-Faktor yang diduga Mempengaruhi Kemiskinan.....	29
<b>BAB 3 METODOLOGI PENELITIAN .....</b>		<b>33</b>
3.1	Sumber Data .....	33
3.2	Kerangka Konseptual Penelitian.....	34
3.3	Variabel Penelitian.....	34
3.4	Struktur Data Penelitian.....	37
3.5	Tahapan Penelitian.....	37
<b>BAB 4 HASIL DAN PEMBAHASAN .....</b>		<b>45</b>
4.1	Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian .....	45
4.1.1	Estimator Kurva Regresi untuk Fungsi <i>Spline Truncated</i> Linear Multivariabel .....	45
4.1.2	Estimator Kurva Regresi untuk Fungsi Kernel Gaussian .....	49
4.1.3	Estimator Campuran Fungsi <i>Spline Truncated</i> Multivariabel dan Kernel Gaussian.....	51
4.1.4	<i>Cross-Validation (CV)</i> .....	54
4.1.5	<i>Generalized Cross-Validation (GCV)</i> .....	55
4.1.6	<i>Unbiased Risk (UBR)</i> .....	56
4.2	Kajian Simulasi.....	57
4.3	Aplikasi pada Data Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan.....	63
1.	Statistika Deskriptif Variabel Penelitian.....	64
2.	Deteksi Multikolinearitas.....	70
3.	Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan menggunakan Metode CV.....	71
4.	Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan menggunakan Metode GCV.....	76
5.	Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan menggunakan Metode UBR.....	80



6.	Ringkasan Hasil Metode CV, GCV dan UBR dalam Pemodelan Regresi Nonparametrik Menggunakan Estimator Campuran .....	85
7.	Estimasi Parameter dan Interpretasi Model .....	88
4.4	Pembahasan .....	97
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>		<b>99</b>
5.1	Kesimpulan .....	99
5.2	Saran .....	101
<b>DAFTAR PUSTAKA .....</b>		<b>103</b>
<b>LAMPIRAN .....</b>		<b>113</b>
<b>BIOGRAFI PENULIS .....</b>		<b>138</b>

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

	Halaman
<b>Tabel 2.1</b> Jenis dan Fungsi Kernel .....	17
<b>Tabel 3.1</b> Jumlah Kabupaten/Kota di Pulau Kalimantan Berdasarkan Provinsi .....	33
<b>Tabel 3.2</b> Variabel Penelitian.....	35
<b>Tabel 3.3</b> Struktur Data Penelitian.....	37
<b>Tabel 4.1</b> Hasil Simulasi dengan Metode CV.....	59
<b>Tabel 4.2</b> Hasil Simulasi dengan Metode GCV .....	60
<b>Tabel 4.3</b> Hasil Simulasi dengan Metode UBR .....	60
<b>Tabel 4.4</b> Statistika Deskriptif .....	64
<b>Tabel 4.5</b> Komponen Estimator <i>Spline Truncated</i> dan Kernel .....	70
<b>Tabel 4.6</b> Deteksi Multikolinearitas dari Variabel Prediktor.....	71
<b>Tabel 4.7</b> Nilai CV untuk Estimator Campuran dengan Satu Titik Knot .....	71
<b>Tabel 4.8</b> Nilai CV untuk Estimator Campuran dengan Dua Titik Knot.....	73
<b>Tabel 4.9</b> Nilai CV untuk Estimator Campuran dengan Tiga Titik Knot.....	74
<b>Tabel 4.10</b> Ringkasan dengan Metode CV .....	75
<b>Tabel 4.11</b> Nilai GCV untuk Estimator Campuran dengan Satu Titik Knot ...	76
<b>Tabel 4.12</b> Nilai GCV untuk Estimator Campuran dengan Dua Titik Knot....	77
<b>Tabel 4.13</b> Nilai GCV untuk Estimator Campuran dengan Tiga Titik Knot ...	79
<b>Tabel 4.14</b> Ringkasan dengan Metode GCV .....	80
<b>Tabel 4.15</b> Nilai UBR untuk Estimator Campuran dengan Satu Titik Knot ...	81
<b>Tabel 4.16</b> Nilai UBR untuk Estimator Campuran dengan Dua Titik Knot....	82
<b>Tabel 4.17</b> Nilai UBR untuk Estimator Campuran dengan Tiga Titik Knot ...	83
<b>Tabel 4.18</b> Ringkasan dengan Metode UBR.....	85
<b>Tabel 4.19</b> Perbandingan Metode CV, GCV dan UBR .....	85
<b>Tabel 4.20</b> Uji Normalitas dengan Kolmogorov-Smirnov Test.....	86
<b>Tabel 4.21</b> Estimasi Parameter Model Regresi Nonparametrik.....	88

*(Halaman ini sengaja dikosongkan)*

## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Grafik Dari Beberapa Model Regresi Parametrik .....	12
<b>Gambar 2.2</b> Ilustrasi Pola Perubahan Data Berdasarkan Titik Knot .....	21
<b>Gambar 2.3</b> Ilustrasi Pemilihan <i>Bandwidth</i> .....	22
<b>Gambar 3.1</b> Peta Pulau Kalimantan .....	33
<b>Gambar 3.2</b> Variabel yang diduga Mempengaruhi Kemiskinan .....	34
<b>Gambar 3.3</b> Langkah Analisis untuk Tujuan Pertama .....	42
<b>Gambar 3.3</b> Langkah Analisis untuk Tujuan Kedua .....	43
<b>Gambar 3.4</b> Langkah Analisis untuk Tujuan Ketiga .....	43
<b>Gambar 4.1</b> Diagram Pencar antara Variabel Prediktor dengan Variabel Respon Data Simulasi.....	58
<b>Gambar 4.2</b> Ilustrasi Dampak dari Varians pada Pola Data.....	62
<b>Gambar 4.3</b> Diagram Pencar antara $x_1$ dengan $y$ .....	66
<b>Gambar 4.4</b> Diagram Pencar antara $x_2$ dengan $y$ .....	67
<b>Gambar 4.5</b> Diagram Pencar antara $x_3$ dengan $y$ .....	67
<b>Gambar 4.6</b> Diagram Pencar antara $x_4$ dengan $y$ .....	68
<b>Gambar 4.7</b> Diagram Pencar antara $x_5$ dengan $y$ .....	68
<b>Gambar 4.8</b> Diagram Pencar antara $x_6$ dengan $y$ .....	69
<b>Gambar 4.9</b> Diagram Pencar antara $t_1$ dengan $y$ .....	69
<b>Gambar 4.10</b> Visualisasi Nilai $y$ dengan $\hat{y}$ dengan Metode CV .....	87
<b>Gambar 4.11</b> Visualisasi Nilai $y$ dengan $\hat{y}$ dengan Metode GCV .....	87
<b>Gambar 4.12</b> Visualisasi Nilai $y$ dengan $\hat{y}$ dengan Metode UBR.....	87



*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1</b> Data Penelitian.....	114
<b>Lampiran 2</b> <i>Syntax</i> R Studi Simulasi Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian dengan Metode <i>Cross-Validation</i> (CV).....	116
<b>Lampiran 3</b> <i>Syntax</i> R Studi Simulasi Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian dengan Metode <i>Generalized Cross-Validation</i> (GCV).....	119
<b>Lampiran 4</b> <i>Syntax</i> R Studi Simulasi Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian dengan Metode <i>Unbiased Risk</i> (UBR) .....	122
<b>Lampiran 5</b> <i>Syntax</i> R Data Aplikasi pada Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian dengan Metode <i>Cross-Validation</i> (CV) .....	125
<b>Lampiran 6</b> <i>Syntax</i> R Data Aplikasi pada Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian dengan Metode <i>Generalized Cross-Validation</i> (GCV) .....	129
<b>Lampiran 7</b> <i>Syntax</i> R Data Aplikasi pada Model Regresi Nonparametrik Estimator Campuran <i>Spline Truncated</i> dan Kernel Gaussian dengan Metode <i>Unbiased Risk</i> (UBR).....	133
<b>Lampiran 8</b> Estimasi Variabel Respon ( $\hat{y}$ ) dari Model Regresi Nonparametrik Estimator Campuran Terbaik (GCV 3 Titik Knot).....	137

*(Halaman ini sengaja dikosongkan)*

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Beberapa masalah didunia nyata, seperti misalnya terdapat dua atau lebih variabel yang hubungannya tidak dapat dipisahkan, akan menarik apabila diselidiki sifat hubungannya. Metode yang sering kali digunakan adalah analisis regresi. Analisis regresi adalah salah satu metode statistika yang digunakan untuk mengetahui pola hubungan antara satu atau lebih variabel. Variabel yang digunakan dalam analisis regresi terdiri atas variabel respon dan variabel prediktor. Identifikasi awal untuk memeriksa adanya pola hubungan adalah dengan menggunakan diagram pencar (*scatter plot*). Budiantara (2009) dalam penelitiannya menyebutkan bahwa terdapat hal lain yang perlu diperiksa sebelum melakukan pemodelan statistika dengan analisis regresi, yaitu menyelidiki apakah variabel-variabel tersebut secara rasional saling berkorelasi atau tidak. Apabila secara rasional berkorelasi, maka dapat dilakukan pemodelan statistika dengan menggunakan analisis regresi. Tujuan utama dalam melakukan analisis regresi adalah untuk mencari bentuk estimasi kurva regresi.

Terdapat beberapa pendekatan yang dikembangkan oleh para peneliti dalam pemodelan statistika menggunakan analisis regresi, yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik. Regresi parametrik yaitu regresi yang dimana bentuk pola kurva regresinya diketahui, sedangkan regresi nonparametrik yaitu regresi yang dimana bentuk pola kurva regresinya tidak diketahui. Sementara, apabila sebagian bentuk pola kurva regresinya diketahui dan sebagian lagi tidak diketahui maka analisis regresi yang digunakan adalah regresi semiparametrik. Seringkali dalam prakteknya, bentuk pola hubungan antara variabel respon dengan variabel prediktor tidak diketahui. Pada keadaan seperti ini, tentunya model regresi parametrik kurang cocok apabila digunakan, maka pendekatan yang disarankan adalah regresi nonparametrik. Pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva regresi tertentu, sehingga akan memberikan fleksibilitas

yang tinggi, dimana diharapkan data menyesuaikan sendiri bentuk estimasi kurva regresi tanpa dipengaruhi oleh subyektifitas peneliti (Eubank, 1999).

Pada pemodelan regresi nonparametrik, terdapat beberapa estimator yang dikembangkan oleh para peneliti diantaranya adalah Kernel (Nadaraya, 1989; Hardle, 1994; Eubank, 1999; Sukarsa & Srinadi, 2012; Klemela, 2014; Astuti, dkk., 2018), *Spline* (Wahba, 1990; Eubank, 1999; Lin & Zhang, 2006; Budiantara, 2009; Tripena, 2011; Fatmawati, dkk., 2019; Rahim, dkk., 2019), dan deret Fourier (Bilodeau, 1992; Eubank, 1999; Morton & Silverberg, 1999; Nurjanah, dkk., 2015; Prahutama, dkk., 2018). Diantara beberapa metode-metode tersebut, regresi nonparametrik dengan pendekatan *spline* dan kernel merupakan metode estimator yang biasanya sering digunakan pada penelitian.

*Spline* merupakan model yang memiliki interpretasi statistik dan interpretasi secara visual yang sangat khusus dan sangat baik (Budiantara, 2011). Eubank (1999) menjelaskan bahwa estimator dengan menggunakan *spline* memiliki fleksibilitas yang tinggi. *Spline* juga memiliki kemampuan yang sangat baik dalam menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Eubank, 1999; Budiantara, 2011). Eubank (1999) menjelaskan bahwa dengan menggunakan titik knot, kurva dengan fungsi *spline* mampu mengatasi permasalahan pola data yang menunjukkan naik turunnya perubahan pola data yang tajam sehingga dapat dihasilkan estimasi kurva yang relatif mulus. Montoya, dkk. (2014) menjelaskan bahwa bentuk kurva regresi nonparametrik dengan pendekatan fungsi *spline* sangat dipengaruhi oleh banyaknya titik knot dan lokasi dimana titik knot berada.

*Spline* pertama kali diperkenalkan oleh Whitaker pada Tahun 1923 sebagai pendekatan pola data. Pendekatan *spline* mempunyai suatu basis fungsi, salah satunya adalah *spline truncated*. *Spline truncated* merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Salah satu kelebihan *spline* seperti yang telah dijelaskan sebelumnya adalah bersifat fleksibel, artinya fungsi ini cenderung mencari sendiri bentuk estimasi data kemanapun pola data tersebut bergerak. *Spline* mempunyai kemampuan yang sangat baik untuk di generalisasikan pada pemodelan statistika yang bersifat kompleks dan rumit. Beberapa penelitian sebelumnya yang menggunakan model regresi nonparametrik *spline truncated*



yaitu Tripena (2011) dengan melakukan penentuan model regresi *spline* terbaik menggunakan metode *Generalized Cross-Validation* (GCV). Budiantara, dkk. (2012) memodelkan kemiskinan di Indonesia menggunakan pendekatan *spline truncated*. Setiawan, dkk. (2017) menggunakan model regresi nonparametrik *spline truncated* pada data Indeks Pembangunan Gender (IPG). Selain itu, Saputro, dkk. (2018) menggunakan model regresi nonparametrik *spline truncated* pada data Indeks Pembangunan Manusia di Indonesia.

Selain *spline*, pendekatan regresi nonparametrik dengan fungsi kernel juga mendapat banyak perhatian dari para peneliti. Kernel memiliki bentuk yang lebih fleksibel, baik dalam memodelkan data yang tidak memiliki pola tertentu, selain itu perhitungan matematisnya mudah (Hardle, 1994). Estimator kernel sangat bergantung pada pemilihan fungsi kernel dan *bandwidth* yang digunakan. Hardle (1990) dalam penelitian Suparti dan Prahutama (2016) menjelaskan bahwa terdapat beberapa jenis fungsi kernel yaitu Uniform, Segitiga, Epanechnikov, Kuadrat, Triweight, Cosinus dan Gaussian. Hastie dan Tibshirani (1990) menjelaskan bahwa fungsi kernel Gaussian adalah fungsi kernel yang sering digunakan dalam penelitian, di karenakan memiliki sifat yang fleksibel dan *smooth*. Beberapa penelitian sebelumnya yang menggunakan model regresi nonparametrik dengan pendekatan kernel yaitu Kayri dan Zirhlioglu (2009), penelitian menggunakan pemodelan dengan estimator kernel. Du, dkk. (2013) meneliti tentang model regresi kernel multivariabel prediktor. Aljuhani dan Alturk (2014) melakukan penelitian yaitu memodifikasi estimasi regresi kernel adaptif Nadaraya-Watson. Aydin, dkk. (2016) melakukan penelitian tentang pemilihan *bandwidth* optimal dalam regresi nonparametrik kernel.

Model regresi nonparametrik yang saat ini berkembang, terbatas hanya pada satu bentuk estimator untuk keseluruhan variabel prediktor yang digunakan. Hal ini terjadi, dikarenakan adanya asumsi atau *statement* dari peneliti bahwa masing-masing dari prediktor tersebut mempunyai pola data yang sama. Namun pada kenyataannya, ketika memodelkan variabel prediktor dengan variabel respon, hal yang terjadi adalah setiap prediktor kemungkinan memiliki pola yang berbeda. Apabila dipaksakan hanya menggunakan satu bentuk estimator, maka akan menghasilkan bentuk estimasi yang tidak sesuai dengan pola data dan nantinya akan

berdampak terhadap akurasi yang buruk. Budiantara, dkk. (2015) selanjutnya mengembangkan estimator campuran *spline* dan kernel, namun penelitian ini terbatas hanya pada penggunaan satu prediktor untuk setiap komponen *spline* dan kernel. Ratnasari, dkk. (2016) kemudian menggeneralisasikan penelitian Budiantara, dkk. (2015) dimana komponen estimator *spline* dan kernel mengandung lebih dari satu prediktor. Purnomo (2016) juga melakukan penelitian mengenai estimator campuran *spline* dan kernel dalam model regresi nonparametrik dengan menggunakan metode estimasi *Ordinary Least Squares* (OLS). Hidayat, dkk. (2019) melakukan penelitian menggunakan Kernel-Spline estimator pada model aditif regresi nonparametrik.

Pada regresi nonparametrik *spline truncated* menentukan titik knot optimal menjadi hal yang sangat penting dan krusial. Hal yang sama berlaku pula pada regresi kernel, dimana hal yang harus diperhatikan adalah pemilihan *bandwidth*. Maka menjadi permasalahan tersendiri apabila menggunakan model estimator campuran, khususnya mengkombinasikan estimator *spline* dengan kernel sedemikian sehingga harus menentukan titik knot dan *bandwidth* yang optimal.

Penentuan titik knot dan *bandwidth* dalam estimator campuran regresi nonparametrik *spline truncated* dan kernel akan sangat mempengaruhi kurva regresi yang akan dibentuk. Untuk memilih titik knot dan parameter *bandwidth* yang optimal dalam regresi nonparametrik dapat menggunakan berbagai metode yang telah dikembangkan oleh peneliti, sebagai contoh *Cross-Validation* (CV), *Generalized Cross-Validation* (GCV), *Unbiased Risk* (UBR), dan *Generalized Maximum Likelihood* (GML). Pada penelitian ini hanya akan dibahas mengenai metode CV, GCV dan UBR.

*Cross-Validation* (CV) adalah suatu metode yang diperkenalkan oleh Craven dan Wahba (1979). Metode CV biasa disebut sebagai metode hapus-satu, yaitu suatu metode yang bertujuan untuk meminimumkan jumlah kuadrat dari *error* prediksi untuk variabel respon (Andrews, 1991). Metode CV mengasumsikan bahwa fungsi diperoleh tanpa pengamatan ke-*i* dari data, yang kemudian estimator diperoleh dari suatu sub-sampel yang diambil dari data asli (Eubank, 1999). Kelebihan metode CV adalah nilai estimasi pada CV dihitung dari regresi lokal yang menghilangkan pengamatan data ke-*i* (Fox, 2008). Metode CV memberikan

bobot yang berbeda-beda pada masing-masing pengamatan sesuai kontribusinya (Devi, dkk., 2018).

*Unbiased Risk* (UBR) adalah suatu metode yang diperkenalkan oleh Wang (1997). UBR merupakan suatu metode pemilihan parameter penghalus yang mensyaratkan nilai estimasi dari varians *error* diketahui, yang artinya metode UBR memperhatikan varians. Kelebihan dari metode UBR adalah baik digunakan untuk memilih parameter penghalus untuk data non-Gaussian, sedangkan Wahba (1990) menjelaskan bahwa metode GCV akan lebih baik jika digunakan pada data Gaussian (Gu, dkk., 1992; Wang, 1997).

*Generalized Cross-Validation* (GCV) adalah suatu metode yang dikembangkan oleh Wahba (1990). GCV merupakan pengembangan dari metode CV. Metode GCV merupakan metode CV yang terboboti, yaitu dengan memberikan bobot yang sama pada setiap pengamatan (Devi, dkk., 2018). Nilai GCV diperoleh dengan menjumlahkan residual-residual kuadrat yang telah terkoreksi (Pratiwi, 2017). Kelebihan yang dimiliki metode GCV adalah sederhana dan efisien dalam perhitungan, invariant terhadap transformasi, optimal secara asimtotik dan tidak memerlukan informasi terhadap  $\sigma^2$  (Wahba, 1990).

Pengkajian mengenai metode CV, GCV dan UBR pernah dilakukan oleh Wahba (1990), Eubank (1999), Wang (1997) yang terbatas pada estimator *spline smoothing* pada data univariabel. Budiantara (2000) melakukan kajian mengenai metode CV, GCV dan GML pada model regresi nonparametrik *spline*. Selanjutnya Sari, dkk., (2016) membandingkan metode GCV dan UBR, dan diperoleh bahwa metode GCV lebih baik dibandingkan metode UBR. Devi, dkk., (2018) kemudian membandingkan metode UBR dan CV dalam proses pemilihan titik knot optimal pada model regresi nonparametrik *spline truncated* multivariabel, diperoleh bahwa metode CV lebih baik dibandingkan metode UBR.

Hal yang mendasari penelitian ini adalah adanya tiga kriteria ketika melakukan pemodelan regresi nonparametrik khususnya menggunakan estimator campuran *spline truncated* dan kernel Gaussian, yaitu (1) menentukan banyaknya titik knot, (2) lokasi dimana titik knot itu berada dan (3) menentukan *bandwidth* yang tepat. Berdasarkan uraian tersebut dan mempertimbangkan kelebihan setiap

metode CV, GCV dan UBR pada model estimator tunggal, maka dalam penelitian ini akan dilakukan kajian lanjutan pada model estimator campuran regresi nonparametrik. Fokus penelitian ini ditekankan pada pemilihan titik knot dan *bandwidth* yang optimal pada estimator campuran *spline truncated* dan kernel Gaussian.

Model regresi nonparametrik, dapat diterapkan di berbagai bidang keilmuan, salah satunya adalah bidang ekonomi. Pada bidang ini, model regresi nonparametrik dapat diterapkan untuk menganalisis hal-hal yang berkaitan dengan kemiskinan. Kemiskinan secara singkat seringkali dipahami sebagai rendahnya tingkat kesejahteraan. Masalah kemiskinan merupakan masalah yang sangat serius yang dihadapi oleh negara-negara didunia terutama negara-negara berkembang termasuk Indonesia. Badan Pusat Statistik (BPS) mendefinisikan kemiskinan sebagai ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak, selanjutnya kemiskinan merupakan sebuah kondisi yang berada dibawah garis nilai standar kebutuhan minimum, baik untuk makanan dan non-makanan yang disebut garis kemiskinan (*poverty line*).

Berdasarkan data BPS (2020), pada Maret 2020 persentase penduduk miskin di Indonesia adalah 9,78%, mengalami peningkatan sebesar 0,56% poin terhadap September 2019 dan meningkat 0,37 persen poin terhadap Maret 2019. Jumlah penduduk miskin pada Maret 2020 sebesar 26,42 juta orang, dimana meningkat sebesar 1,63 juta orang terhadap September 2019. Garis kemiskinan pada Maret 2020 tercatat sebesar Rp. 454.652, -/Kapita/Bulan.

Penelitian mengenai kemiskinan menggunakan pendekatan model regresi nonparametrik sebelumnya pernah dilakukan oleh Budiantara, dkk. (2011) yang menganalisis pola hubungan kemiskinan dan pengangguran di Indonesia dengan menggunakan pendekatan *spline* Bayesian. Mubarak dan Budiantara (2012) menggunakan model regresi regresi nonparametrik *spline truncated* multivariabel dimana salah satu variabel yang digunakan adalah data persentase penduduk miskin di Jawa Timur. Merdekawati dan Budiantara (2013) memodelkan kemiskinan di Provinsi Jawa Tengah menggunakan pendekatan *spline truncated* multivariabel. Budiantara, dkk. (2015) melakukan pemodelan persentase penduduk miskin di Indonesia dengan menggunakan pendekatan regresi nonparametrik *spline*.

Ratnasari, dkk. (2015) melakukan analisis tentang kemiskinan di Provinsi Papua dengan membandingkan estimator *spline truncated* dan deret Fourier.

Berdasarkan uraian diatas, masih belum terdapat penelitian yang mengkaji pemilihan titik knot dan *bandwidth* optimal pada model estimator campuran regresi nonparametrik, khususnya dengan mengkombinasikan *spline truncated* dan kernel Gaussian. Oleh karena itu, peneliti tertarik melakukan penelitian untuk membandingkan hasil pemilihan titik knot dan *bandwidth* yang optimal dengan metode *Cross-Validation* (CV), *Generalized Cross-Validation* (GCV) dan *Unbiased Risk* (UBR) pada model regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian. Pada penelitian ini juga akan dilakukan kajian simulasi untuk melihat kebaikan metode CV, GCV dan UBR dalam menduga titik knot dan *bandwidth* yang optimal. Studi kasus pada penelitian ini yaitu persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019. Diharapkan hasil penelitian ini dapat bermanfaat dalam pengembangan ilmu statistika khususnya pada bidang regresi nonparametrik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, pemilihan titik knot dan *bandwidth* yang optimal merupakan hal yang sangat penting dan krusial dalam proses pemodelan regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian. Selanjutnya pendekatan regresi nonparametrik tersebut akan digunakan untuk memodelkan data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan, karena diduga memiliki bentuk kurva regresi yang dimana variabel respon dengan variabel prediktor tidak diketahui polanya dan setiap prediktor memiliki pola yang berbeda, sehingga perlu menggunakan model estimator campuran.

Rumusan masalah yang dapat diuraikan pada penelitian ini adalah sebagai berikut:

1. Bagaimana bentuk estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian?

2. Bagaimana hasil perbandingan kinerja metode pemilihan titik knot dan *bandwidth* menggunakan CV, GCV dan UBR yang diterapkan pada data simulasi?
3. Bagaimana hasil perbandingan metode pemilihan titik knot dan *bandwidth* yang optimal menggunakan CV, GCV dan UBR yang kemudian diaplikasikan pada data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019?

### 1.3 Tujuan Penelitian

Sesuai dengan permasalahan yang telah dipaparkan sebelumnya, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengkaji bentuk estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian.
2. Melakukan studi simulasi pada model estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian untuk membandingkan kinerja metode CV, GCV dan UBR dalam menduga titik knot dan *bandwidth* yang optimal.
3. Membandingkan metode pemilihan titik knot dan *bandwidth* yang optimal menggunakan CV, GCV dan UBR yang kemudian diaplikasikan pada data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019.

### 1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan peneliti dari permasalahan mengenai persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan dalam penelitian ini adalah:

1. Menambah wawasan keilmuan kepada peneliti khususnya dan masyarakat pada umumnya tentang regresi nonparametrik *spline truncated*, regresi kernel dan pengembangan menggunakan estimator campuran.
2. Menghasilkan kajian teoritis tentang bentuk estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian sehingga dapat dijadikan tambahan referensi dalam melakukan analisis data.

3. Memberikan informasi secara lebih rinci mengenai persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan sehingga dapat dijadikan masukan guna membantu pemerintah daerah dan pemerintah pusat dalam menentukan kebijakan terkait pengentasan kemiskinan.

### **1.5 Batasan Masalah**

Batasan masalah yang digunakan agar penelitian dapat berjalan secara terarah serta dapat mencapai tujuan yang diinginkan adalah sebagai berikut:

1. Model regresi yang digunakan adalah regresi nonparametrik dengan estimator campuran yaitu *spline truncated* multivariabel dengan fungsi linier dan regresi kernel dengan fungsi Gaussian.
2. Banyaknya titik knot yang digunakan pada proses pemodelan regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian sebanyak 1, 2 hingga 3 titik knot.
3. Metode estimasi yang digunakan dalam penelitian ini adalah *Ordinary Least Square* (OLS).
4. Pada bagian kajian simulasi, *error* yang dibangkitkan mengikuti distribusi Normal  $(0, \sigma^2)$ .
5. Data yang digunakan dalam penelitian ini adalah persentase penduduk miskin di Pulau Kalimantan Tahun 2019 yang merupakan hasil publikasi BPS Tahun 2020. Observasi awal pengamatan pada penelitian ini adalah 56 Kabupaten/Kota yang ada di Pulau Kalimantan. Yang selanjutnya untuk Kota Bontang, Banjarmasin dan Pontianak tidak digunakan dalam observasi penelitian karena merupakan data pencilan, sehingga observasi pengamatan pada penelitian ini hanya tersisa 53 Kabupaten/Kota.

*(Halaman ini sengaja dikosongkan)*



## BAB 2

### TINJAUAN PUSTAKA

Pada bagian ini akan dibahas mengenai beberapa landasan teori yang terkait guna menyelesaikan permasalahan pada rumusan masalah penelitian mengenai persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan yang meliputi analisis regresi, regresi nonparametrik, regresi *spline truncated*, regresi kernel, estimator campuran, pemilihan titik knot dan *bandwidth*, kriteria pemilihan model terbaik dan definisi dari penduduk miskin serta faktor-faktor yang diduga mempengaruhinya.

#### 2.1 Analisis Regresi

Analisis regresi merupakan metode statistika yang dapat digunakan untuk mengetahui pola hubungan antar variabel prediktor dan variabel respon (Drapper dan Smith, 1992). Secara umum, model analisis regresi dengan pasangan data  $(x_i, y_i)$  dapat dituliskan sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i; i = 1, 2, \dots, n \quad (2.1)$$

dimana

- $y_i$  : variabel respon
- $x_i$  : variabel prediktor
- $f(x_i)$  : kurva regresi yang akan di hampiri
- $\varepsilon_i$  : *error* acak yang diasumsikan identik, *independent* dan berdistribusi normal dengan *mean* nol dan varians  $\sigma^2$  atau  $\varepsilon_i \sim N(0, \sigma^2)$

Tujuan dari analisis regresi adalah untuk memperkirakan atau memprediksi suatu nilai dari variabel respon apabila variabel prediktor ditetapkan nilainya, dengan kata lain adalah mencari bentuk estimasi parameter yang sesuai dengan bentuk dari kurva regresi. Jika bentuk kurva regresi diketahui, maka dapat digunakan regresi parametrik, sedangkan jika bentuk kurva regresi tidak diketahui dan tidak ada informasi yang lengkap dan detail sebelumnya maka dapat menggunakan regresi nonparametrik. Adapun regresi semiparametrik dapat

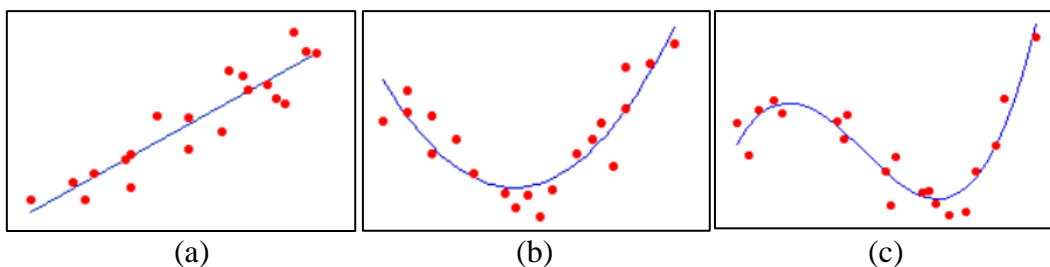
digunakan ketika pola hubungan antara variabel prediktor dan variabel respon merupakan kombinasi antara parametrik dan nonparametrik (Budiantara, 2019).

## 2.2 Regresi Parametrik dan Nonparametrik

Salah satu tahapan awal dalam analisis regresi adalah melakukan pendugaan bentuk kurva regresi. Pendugaan bentuk kurva regresi digunakan untuk menjelaskan hubungan antara variabel respon dengan variabel prediktor.

Model regresi umumnya dibagi menjadi 3 bagian, yaitu model regresi parametrik, model regresi nonparametrik dan model regresi semiparametrik. Bentuk dari model regresi sangat tergantung pada kurva  $f(x_i)$  pada Persamaan (2.1). Apabila bentuk kurva  $f(x_i)$  diketahui, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi parametrik. Namun apabila bentuk kurva  $f(x_i)$  tidak diketahui atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola datanya, maka pendekatan model regresi yang digunakan adalah pendekatan model regresi nonparametrik (Budiantara, 2009). Dalam beberapa kasus, sebagian bentuk pola data diketahui, sedangkan sebagian lainnya tidak diketahui bentuk pola datanya. Pada kasus ini, pendekatan model regresi yang disarankan adalah pendekatan model regresi semiparametrik (Wahba, 1990).

Pendekatan pendugaan yang sering dijumpai adalah pendekatan parametrik. Asumsi yang mendasari pendekatan regresi parametrik adalah bahwa bentuk dari kurva regresi dapat digambarkan dalam pola tertentu (Hardle, 1994). Jika kurva regresi mengikuti bentuk yang telah diketahui sebelumnya seperti linier, kuadratik, kubik, eksponensial dan lain sebagainya, maka dapat digunakan pendekatan regresi parametrik. Beberapa grafik ditampilkan sebagai berikut: (Budiantara, 2019)



**Gambar 2.1** Grafik dari beberapa model regresi parametrik (a) linier; (b) kuadratik; (c) kubik

Pendekatan model regresi parametrik memiliki sifat yang sangat baik dari sisi statistika inferensi yaitu: sederhana, memiliki estimator yang tidak bias, efisien dan konsisten serta mudah untuk diinterpretasikan, yang sangat jarang dimiliki oleh pendekatan regresi lainnya seperti regresi nonparametrik dan regresi semiparametrik (Budiantara, 2009). Estimasi parameter dari model diperoleh berdasarkan berbagai metode penaksir yang telah diketahui dalam ilmu statistika, seperti *least square*, *maximum likelihood estimation*, dan *Bayesian*.

Hardle (1994), menyatakan bahwa pendekatan regresi nonparametrik mulai dikenal dikisaran abad-ke 19 tepatnya pada Tahun 1857. Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan variabel respon yang tidak diketahui bentuk kurva regresinya dan tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola datanya (Eubank, 1999). Berdasarkan hal itu, maka secara visual pola data yang diberikan oleh variabel prediktor dan variabel respon tidak mempunyai pola yang mengikuti pola tertentu. Dalam pandangan regresi nonparametrik, data akan mencari sendiri bentuk estimasi dari kurva regresinya tanpa harus dipengaruhi oleh faktor subyektifitas peneliti. Ini berarti pendekatan model regresi nonparametrik memiliki fleksibilitas yang tinggi (sangat fleksibel) dan obyektif (Eubank, 1999).

Secara umum model regresi nonparametrik memiliki bentuk fungsi seperti bentuk regresi pada umumnya seperti pada Persamaan (2.1). Regresi nonparametrik disebut regresi nonparametrik univariabel jika hanya terdiri dari satu variabel respon dan satu variabel prediktor. Sedangkan jika dalam regresi nonparametrik terdapat satu variabel respon dan lebih dari satu variabel prediktor, maka disebut regresi nonparametrik multivariabel (Budiantara, 2004).

### **2.3 Regresi Nonparametrik *Spline Truncated***

Pendekatan regresi nonparametrik *spline truncated* digunakan jika kurva regresi antara variabel respon dengan variabel prediktor tidak membentuk suatu pola atau tidak terdapat informasi masa lalu yang lengkap mengenai pola dan struktur data. Dalam prakteknya pada kasus-kasus riil, pengamatan-pengamatan yang dikaji tidak selalu memenuhi asumsi yang mendasari pendekatan parametrik, sedemikian sehingga kerap kali dibutuhkan suatu teknik inferensi dengan validitas

yang tidak bergantung pada asumsi-asumsi yang kaku. Menurut Eubank (1999), pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi dan tidak memiliki asumsi mengenai bentuk pola hubungan antara variabel respon dengan variabel prediktor. Secara umum, model regresi nonparametrik seperti pada Persamaan (2.1). Pendekatan regresi nonparametrik yang sangat terkenal dan populer adalah *spline*.

*Spline* merupakan model polinomial yang tersegmen. Mengingat adanya kelemahan-kelemahan yang dimiliki oleh fungsi polinomial, dimana salah satunya bersifat global, maka dikembangkan suatu fungsi yang dapat mengatasi kelemahan-kelemahan polinomial, yaitu fungsi *spline truncated*. Fungsi *spline truncated* adalah suatu fungsi yang masih mempertahankan sifat dari fungsi polinomial, yang dirancang dengan cara memodifikasi fungsi polinomial. Fungsi *spline truncated* derajat  $m$  dengan  $\phi$  merupakan titik-titik knot yaitu  $K_1, K_2, \dots, K_r$  diberikan oleh persamaan:

$$f(x_i) = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m \quad (2.2)$$

*Spline truncated* memiliki titik knot yang merupakan titik perpaduan yang menunjukkan perubahan perilaku kurva pada selang yang berbeda (Hardle, 1990). Salah satu kelebihan *spline truncated* adalah model ini cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena didalam fungsi *spline truncated* terdapat titik-titik knot (Budiantara, 2019).

Misalkan diberikan data berpasangan  $x_i$  dan  $y_i$ , dimana  $i = 1, 2, \dots, n$  yang mengikuti model regresi nonparametrik. Persamaan (2.2) apabila selanjutnya disubstitusikan ke dalam Persamaan (2.1) akan diperoleh persamaan regresi nonparametrik *spline truncated* sebagai berikut:

$$y_i = \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m + \varepsilon_i; i = 1, 2, \dots, n \quad (2.3)$$

dengan fungsi *truncated* diberikan oleh:

$$(x_i - K_k)_+^m = \begin{cases} (x_i - K_k)^m & x \geq K_k \\ 0 & x < K_k \end{cases} \quad (2.4)$$

Berdasarkan Persamaan (2.4) dapat disajikan untuk setiap  $i = 1, 2, \dots, n$  sehingga dijabarkan sebagai berikut:

$$\begin{aligned}
i = 1 &\rightarrow y_1 = \sum_{j=0}^m \beta_j x_1^j + \sum_{k=1}^r \beta_{m+k} (x_1 - K_k)_+^m + \varepsilon_1 \\
i = 2 &\rightarrow y_2 = \sum_{j=0}^m \beta_j x_2^j + \sum_{k=1}^r \beta_{m+k} (x_2 - K_k)_+^m + \varepsilon_2 \\
&\vdots \\
i = n &\rightarrow y_n = \sum_{j=0}^m \beta_j x_n^j + \sum_{k=1}^r \beta_{m+k} (x_n - K_k)_+^m + \varepsilon_n
\end{aligned} \tag{2.5}$$

Model regresi pada Persamaan (2.5) dapat dituliskan dalam bentuk matriks:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^m & (x_1 - K_1)_+^m & \cdots & (x_1 - K_r)_+^m \\ 1 & x_2^1 & x_2^2 & \cdots & x_2^m & (x_2 - K_1)_+^m & \cdots & (x_2 - K_r)_+^m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & x_n^2 & \cdots & x_n^m & (x_n - K_1)_+^m & \cdots & (x_n - K_r)_+^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{m+r} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.6}$$

Maka dapat dituliskan dalam bentuk matriks menjadi:

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon} \tag{2.7}$$

dimana  $\vec{Y}$  merupakan vektor dari variabel respon berukuran  $(n \times 1)$ ,  $\mathbf{X}$  merupakan matriks berukuran  $n \times (m + r + 1)$ ,  $\vec{\beta}$  adalah vektor parameter koefisien regresi yang akan diestimasi dan berukuran  $(m + r + 1) \times 1$  serta  $\vec{\varepsilon}$  adalah vektor *error* acak berukuran  $(n \times 1)$ . Selanjutnya parameter  $\vec{\beta}$  dapat diestimasi dengan menggunakan *ordinary least squares* (OLS) yaitu metode untuk mendapatkan estimator dengan cara meminimumkan jumlah kuadrat *error*. Adapun *error* dari model regresi pada Persamaan (2.7) diberikan oleh:

$$\varepsilon_i = y_i - \left[ \sum_{j=0}^m \beta_j x_i^j + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m \right] \rightarrow \vec{Y} - \mathbf{X}\vec{\beta} \tag{2.8}$$

Sehingga jumlah kuadrat *error* diberikan oleh:

$$\sum_{i=1}^n \varepsilon_i^2 = \vec{\varepsilon}^T \vec{\varepsilon} \rightarrow (\vec{Y} - \mathbf{X}\vec{\beta})^T (\vec{Y} - \mathbf{X}\vec{\beta}) \tag{2.9}$$

Metode OLS mendapatkan estimator  $\hat{\vec{\beta}}$  melalui penyelesaian optimasi:

$$\min \{Q(\vec{\beta})\} = \min \{(\vec{Y} - \mathbf{X}\vec{\beta})^T (\vec{Y} - \mathbf{X}\vec{\beta})\} \tag{2.10}$$

Untuk menyelesaikan optimasi pada Persamaan (2.10) dapat menggunakan derivatif parsial. Bentuk kuadratik dari  $Q(\vec{\beta})$  dapat diuraikan menjadi:

$$\begin{aligned} Q(\vec{\beta}) &= (\vec{Y} - \mathbf{X}\vec{\beta})^T (\vec{Y} - \mathbf{X}\vec{\beta}) \\ &= (\vec{Y} - \mathbf{X}^T \vec{\beta}^T) (\vec{Y} - \mathbf{X}\vec{\beta}) \\ &= [\vec{Y}^T \vec{Y} - 2\vec{\beta}^T \mathbf{X}^T \vec{Y} + \vec{\beta}^T \mathbf{X}^T \mathbf{X} \vec{\beta}] \end{aligned} \quad (2.11)$$

Kemudian Persamaan (2.11) diturunkan terhadap vektor  $\vec{\beta}$  dan disamadengankan nol sedemikian sehingga diperoleh:

$$\begin{aligned} \frac{\partial(Q(\vec{\beta}))}{\partial \vec{\beta}} &= \frac{\partial [\vec{Y}^T \vec{Y} - 2\vec{\beta}^T \mathbf{X}^T \vec{Y} + \vec{\beta}^T \mathbf{X}^T \mathbf{X} \vec{\beta}]}{\partial \vec{\beta}} = 0 \\ &= -2\mathbf{X}^T \vec{Y} + 2\mathbf{X}^T \mathbf{X} \vec{\beta} = 0 \end{aligned} \quad (2.12)$$

Sehingga didapatkan estimasi parameter nya adalah  $\hat{\vec{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{Y}$ . Estimasi kurva regresi dari  $\hat{f}(x_i)$  diperoleh:

$$\begin{aligned} \hat{f}(x_i) &= \mathbf{X} \hat{\vec{\beta}} \\ &= \mathbf{X} [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{Y}] \\ &= \mathbf{A}(\phi) \vec{Y} \end{aligned} \quad (2.13)$$

Dimana matriks  $\mathbf{A}(\phi)$  diberikan oleh:

$$\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (2.14)$$

## 2.4 Regresi Nonparametrik Kernel

Estimator kernel merupakan estimator linier yang serupa dengan estimator lainnya, namun perbedaannya adalah pada kernel terdapat *bandwidth* (Eubank, 1999). Estimator kernel memiliki kemampuan yang baik dalam memodelkan data yang tidak mempunyai pola tertentu.

Misalkan diberikan data berpasangan  $t_i$  dan  $y_i$ , dimana  $i = 1, 2, \dots, n$  yang mengikuti model regresi nonparametrik. Dengan  $y_i$  adalah variabel respon, sedangkan  $t_i$  adalah variabel prediktor. Selanjutnya hubungan  $t_i$  dan  $y_i$  dapat dimodelkan secara fungsional ke dalam bentuk:

$$y_i = h(t_i) + \varepsilon_i; i = 1, 2, \dots, n \quad (2.15)$$

Dimana kurva regresi dari  $h(t_i)$  merupakan kurva yang tidak diketahui bentuknya. Hardle (1994) menjelaskan bahwa kurva  $h(t_i)$  pada model regresi nonparametrik dapat di hampiri atau di estimasi dengan menggunakan pendekatan kernel yang didasarkan pada fungsi densitas kernel.

Estimator kernel mempunyai beberapa kelebihan diantaranya fleksibel, bentuk matematisnya mudah dan lebih *smooth*. Selanjutnya kurva regresi dari  $h(t_i)$  akan di hampiri oleh fungsi kernel. Estimasi kurva regresi dapat disajikan seperti pada Persamaan (2.16).

$$\begin{aligned} \hat{h}_\alpha(t) &= n^{-1} \sum_{i=1}^n \left[ \frac{K_\alpha(t-t_i)}{n^{-1} \sum_{i=1}^n K_\alpha(t-t_i)} \right] y_i \\ &= n^{-1} \left[ \sum_{i=1}^n W_{\alpha i}(t) \right] y_i \end{aligned} \quad (2.16)$$

dengan:

$$W_{\alpha i}(t) = \frac{K_\alpha(t-t_i)}{n^{-1} \sum_{i=1}^n K_\alpha(t-t_i)}$$

$$K_\alpha(t-t_i) = \frac{1}{\alpha} K\left(\frac{t-t_i}{\alpha}\right)$$

$K$  merupakan fungsi kernel. Menurut Hardle (1990) dalam penelitian Suparti dan Prahutama (2016), fungsi kernel  $K$  terdiri dari beberapa jenis yaitu:

**Tabel 2.1** Jenis dan Fungsi Kernel

No.	Jenis Kernel	Fungsi Kernel
1.	Kernel Uniform	$K(z) = \frac{1}{2}; I_{[-1,1]}(z)$
2.	Kernel Segitiga	$K(z) = (1- z ); I_{[-1,1]}(z)$
3.	Kernel Epanechnikov	$K(z) = \frac{3}{4}(1-z^2); I_{[-1,1]}(z)$
4.	Kernel Kuadrat	$K(z) = \frac{15}{16}(1-z^2)^2; I_{[-1,1]}(z)$

**Tabel 2.1** Jenis dan Fungsi Kernel (lanjutan)

5. Kernel Triweight	$K(z) = \frac{35}{32}(1-z^2)^3; I_{[-1,1]}(z)$
6. Kernel Cosinus	$K(z) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}z\right); I_{[-1,1]}(z)$
7. Kernel Gaussian	$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(-z^2)\right); I_{[-\infty,\infty]}(z)$

Kernel Gaussian cukup sering digunakan dalam banyak penelitian, karena lebih *smooth* dibandingkan fungsi kernel yang lain. Sehingga pada penelitian ini, fungsi kernel yang digunakan adalah kernel Gaussian:

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(-z^2)\right); I_{[-\infty,\infty]}(z) \quad (2.17)$$

Dengan  $z_i = \frac{t-t_i}{\alpha}, i=1,2,\dots,n$ ,  $\alpha$  adalah *bandwidth*,  $t$  adalah suatu nilai yang ditentukan dari variabel prediktor dan  $t_i$  adalah nilai ke- $i$  dari variabel prediktor.

Estimator kernel sangat bergantung pada fungsi kernel  $K$  dan *bandwidth*. Berdasarkan Persamaan (2.16) yang berlaku untuk setiap  $t = t_1, t = t_2, \dots, t = t_n$ , maka dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} \hat{h}_\alpha(t_1) \\ \hat{h}_\alpha(t_2) \\ \vdots \\ \hat{h}_\alpha(t_n) \end{bmatrix} = \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\alpha i}(t_1) y_i \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_n) y_i \end{bmatrix}$$

Dapat dijabarkan sebagai contoh:

$$n^{-1} \sum_{i=1}^n W_{\alpha i}(t_1) y_i = n^{-1} W_{\alpha 1}(t_1) y_1 + n^{-1} W_{\alpha 2}(t_1) y_2 + \dots + n^{-1} W_{\alpha n}(t_1) y_n$$

$$n^{-1} \sum_{i=1}^n W_{\alpha i}(t_2) y_i = n^{-1} W_{\alpha 1}(t_2) y_1 + n^{-1} W_{\alpha 2}(t_2) y_2 + \dots + n^{-1} W_{\alpha n}(t_2) y_n$$

⋮



$$n^{-1} \sum_{i=1}^n W_{\alpha i}(t_n) y_i = n^{-1} W_{\alpha 1}(t_n) y_1 + n^{-1} W_{\alpha 2}(t_n) y_2 + \dots + n^{-1} W_{\alpha n}(t_n) y_n$$

Maka,

$$\begin{aligned} \begin{bmatrix} \hat{h}_\alpha(t_1) \\ \hat{h}_\alpha(t_2) \\ \vdots \\ \hat{h}_\alpha(t_n) \end{bmatrix} &= \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\alpha i}(t_1) y_i \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_n) y_i \end{bmatrix} = \begin{bmatrix} n^{-1} W_{\alpha 1}(t_1) y_1 + n^{-1} W_{\alpha 2}(t_1) y_2 + \dots + n^{-1} W_{\alpha n}(t_1) y_n \\ n^{-1} W_{\alpha 1}(t_2) y_1 + n^{-1} W_{\alpha 2}(t_2) y_2 + \dots + n^{-1} W_{\alpha n}(t_2) y_n \\ \vdots \\ n^{-1} W_{\alpha 1}(t_n) y_1 + n^{-1} W_{\alpha 2}(t_n) y_2 + \dots + n^{-1} W_{\alpha n}(t_n) y_n \end{bmatrix} \\ &= \begin{bmatrix} n^{-1} W_{\alpha 1}(t_1) & n^{-1} W_{\alpha 2}(t_1) & \dots & n^{-1} W_{\alpha n}(t_1) \\ n^{-1} W_{\alpha 1}(t_2) & n^{-1} W_{\alpha 2}(t_2) & \dots & n^{-1} W_{\alpha n}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\alpha 1}(t_n) & n^{-1} W_{\alpha 2}(t_n) & \dots & n^{-1} W_{\alpha n}(t_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \mathbf{D}(\alpha) \vec{Y} \end{aligned} \quad (2.18)$$

Berdasarkan Persamaan (2.18) maka estimator kernel adalah  $\hat{h}_\alpha(t) = \mathbf{D}(\alpha) \vec{Y}$ .

Vektor  $\hat{h}_\alpha(t)$  berukuran  $(n \times 1)$ , vektor  $\vec{Y}$  berukuran  $(n \times 1)$  dan matriks  $\mathbf{D}(\alpha)$  berukuran  $(n \times n)$ .

Pemilihan *bandwidth* yang tepat merupakan hal yang sangat penting dalam pemodelan regresi kernel. Pemodelan regresi kernel sangat tergantung pada nilai *bandwidth* ( $\alpha$ ) yang berguna untuk mengontrol *smoothness* dari kurva estimasi (Aydin, dkk., 2016). *Bandwidth* yang terlalu besar akan menghasilkan kurva estimasi yang sangat *smooth* dan cenderung menuju ke rata-rata dari variabel respon. Namun apabila *bandwidth* yang digunakan terlalu kecil, maka akan menghasilkan kurva estimasi yang kurang *smooth* yaitu hasil estimasi akan menuju ke data. Penggunaan *bandwidth* yang tepat (optimal) akan berdampak terhadap hasil estimasi dengan *mean squared error* (MSE) yang kecil.

## 2.5 Estimator Campuran Regresi Nonparametrik

Regresi campuran nonparametrik adalah suatu model regresi nonparametrik multiprediktor yang kurva regresinya bersifat aditif, dimana kurva regresi tersebut

dihampiri oleh dua atau lebih jenis kurva regresi nonparametrik. Budiantara, dkk. (2015) melakukan penelitian tentang model regresi nonparametrik dengan model aditif yang memiliki dua komponen dalam variabel prediktor. Komponen prediktor pertama, kurva regresinya dihampiri dengan menggunakan regresi *spline truncated*, sedangkan komponen prediktor kedua, kurva regresinya dihampiri dengan regresi kernel.

Misalkan diberikan data berpasangan  $(x_i, t_i, y_i)$  dimana hubungan antara variabel prediktor  $(x_i, t_i)$  dan variabel respon  $(y_i)$  mengikuti model regresi nonparametrik:

$$y_i = \mu(x_i, t_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.19)$$

Bentuk kurva regresi diasumsikan tidak diketahui dan hanya di asumsikan *smooth* dalam arti kontinu dan differensiabel. *Error random*  $\varepsilon_i$  berdistribusi normal dengan rata-rata nol dan  $E(\varepsilon_i^2) = \sigma^2$ .

Selanjutnya kurva regresi  $\mu(x_i, t_i)$  diasumsikan aditif sedemikian sehingga  $\mu(x_i, t_i)$  dapat ditulis ke dalam bentuk:

$$\mu(x_i, t_i) = f(x_i) + h(t_i) \quad (2.20)$$

Dimana  $f(x_i)$  dan  $h(t_i)$  merupakan fungsi-fungsi yang *smooth*.

Permasalahan utama yang harus diselesaikan dalam pemodelan estimator campuran kurva regresi nonparametrik adalah mendapatkan bentuk estimasi dari kurva regresi  $\mu(x_i, t_i)$  yaitu:

$$\hat{\mu}_{\phi, \alpha}(x_i, t_i) = \hat{f}_{\phi, \alpha}(x_i) + \hat{h}_{\alpha}(t_i) \quad (2.21)$$

Dimana parameter  $\alpha$  merupakan *bandwidth* dan  $\phi$  merupakan titik knot. Estimator campuran dengan *spline truncated* dan kernel bisa diperoleh dengan menghampiri kurva regresi  $f(x_i)$  dengan fungsi *spline truncated* derajat  $m$  dan titik knot  $\phi = (K_1, K_2, \dots, K_r)^T$  sedangkan kurva regresi  $h(t_i)$  dengan fungsi kernel Gaussian.

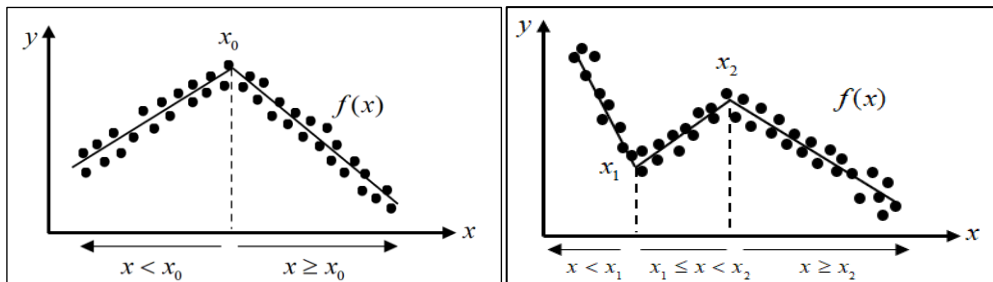
Secara umum model regresi nonparametrik dengan estimator campuran yaitu fungsi *spline truncated* dan fungsi kernel Gaussian dapat disajikan dalam bentuk matriks sebagai berikut:

$$\begin{aligned}
\hat{\mu}_{\phi, \alpha}(x, t) &= \hat{f}_{\phi, \alpha}(x, t) + \hat{h}_{\alpha}(t) \\
&= \mathbf{A}(\phi, \alpha)\vec{Y} + \mathbf{D}(\alpha)\vec{Y} \\
&= [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]\vec{Y} \\
&= \mathbf{B}(\phi, \alpha)\vec{Y}
\end{aligned}
\tag{2.22}$$

Dimana matriks  $\mathbf{B}(\phi, \alpha)$  sangat tergantung pada  $\mathbf{A}(\phi, \alpha)$  yang merupakan komponen *spline truncated* dengan titik knot  $\phi = (K_1, K_2, \dots, K_r)^T$  dan  $\mathbf{D}(\alpha)$  yang merupakan komponen fungsi kernel Gaussian dengan parameter  $\alpha$  yang merupakan *bandwidth*.

## 2.6 Pemilihan Titik Knot dan *Bandwidth* Optimal

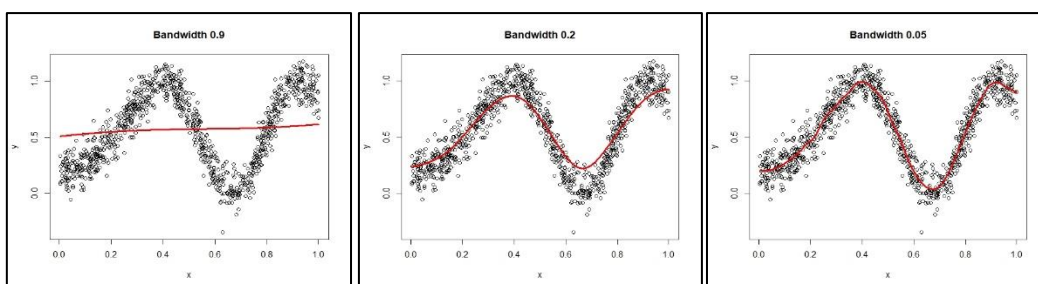
Pada regresi nonparametrik *spline truncated* menentukan titik knot optimal menjadi hal yang sangat penting dan krusial. Titik knot merupakan titik perpaduan yang menunjukkan perubahan perilaku kurva pada selang yang berbeda (Hardle, 1990). Oleh karena itu, dimana ada perilaku data yang polanya berubah, maka di titik tersebutlah letak (lokasi) titik knot. Pola perubahan pola data dapat berbentuk beraneka ragam, disajikan pada Gambar 2.2:



**Gambar 2.2** Ilustrasi pola perubahan data berdasarkan titik knot

- Dari pola data yang awalnya naik, kemudian berubah menjadi pola data turun di titik knotnya.
- Dari pola data yang awalnya turun, kemudian berubah menjadi pola data naik di titik knotnya.
- Dari pola data yang awalnya naik dengan gradien yang cepat (lambat), kemudian berubah menjadi pola data tetap naik di titik knotnya, tetapi dengan gradien yang lambat (cepat) atau pola perubahan yang lainnya.

Disamping lokasi titik knot yang menjadi perhatian dalam *spline truncated*, juga sangat penting untuk memperhatikan berapa banyak titik knot yang harus digunakan. Banyak titik knot biasanya ditentukan dari berapa banyak terjadi perubahan pola perilaku data. Jika terjadi sekali perubahan pola data, maka sebaiknya menggunakan satu titik knot. Jika terjadi dua kali perubahan pola data, maka disarankan menggunakan dua titik knot. Demikian pula dengan seterusnya (Budiantara, 2019).



**Gambar 2.3** Ilustrasi pemilihan *bandwidth*

Hal yang sama berlaku pula pada regresi kernel, dimana hal yang harus diperhatikan adalah pemilihan *bandwidth* ( $\alpha$ ). Menurut Budiantara, dkk. (2015), regresi kernel sangat bergantung pada parameter *bandwidth* dan fungsi kernel yang digunakan. *Bandwidth* merupakan parameter pemulus yang berfungsi untuk mengatur kemulusan kurva yang akan di estimasi, oleh karena itu untuk mendapatkan kurva estimasi yang tepat diperlukan nilai *bandwidth* yang optimal.

Titik knot maupun *bandwidth* yang terlalu kecil akan menghasilkan kurva yang sangat kasar dan cenderung fluktuatif, sebaliknya titik knot maupun *bandwidth* yang terlalu besar akan menghasilkan kurva yang terlalu mulus bahkan sangat mulus, akan tetapi tidak sesuai dengan pola data (Hardle, 1994). Untuk itu menjadi permasalahan tersendiri untuk melakukan pemilihan titik knot  $\phi = (K_1, K_2, \dots, K_r)^T$  dan parameter *bandwidth* ( $\alpha$ ) yang optimal.

Untuk memilih titik knot dan parameter *bandwidth* yang optimal dalam regresi nonparametrik dengan estimator campuran yaitu fungsi *spline truncated* dan fungsi kernel Gaussian, dapat menggunakan berbagai metode yang telah dikembangkan oleh peneliti, misalnya CV, GCV, UBR dan GML (Budiantara, 2019). Dalam penelitian ini, peneliti bertujuan untuk membandingkan hasil pemilihan titik knot dan parameter *bandwidth* yang optimal dengan menggunakan

metode CV, GCV dan UBR pada regresi nonparametrik dengan estimator campuran yaitu fungsi *spline truncated* dan kernel Gaussian.

### 2.8.1 Metode *Cross-Validation* (CV)

Dasar dari metode *Cross-Validation* (CV) adalah untuk memilih titik knot yaitu  $\phi = (K_1, K_2, \dots, K_r)^T$  dan parameter *bandwidth* ( $\alpha$ ) yang meminimumkan CV. Langkah awal menggunakan metode CV adalah memperhatikan bahwa nilainya bergantung secara linier pada data sesuai dengan Persamaan (2.22).

Pemilihan titik knot dan *bandwidth* optimal menggunakan metode CV pada model regresi nonparametrik estimator campuran yang optimal diperoleh dari optimasi:

$$\begin{aligned} CV(\phi_{opt}, \alpha_{opt}) &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{(y_i - \hat{y}_i)}{(1 - [\mathbf{A}_{ii}(\phi, \alpha) + \mathbf{D}_{ii}(\alpha)])} \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{(y_i - \hat{y}_i)}{(1 - [\mathbf{B}_{ii}(\phi, \alpha)])} \right]^2 \end{aligned} \quad (2.23)$$

Dengan kata lain, model regresi nonparametrik dengan estimator campuran yaitu fungsi *spline truncated* dan fungsi kernel Gaussian yang optimal diperoleh dari nilai CV terkecil (Pratiwi, 2017).

### 2.8.2 Metode *Generalized Cross-Validation* (GCV)

Metode *Generalized Cross-Validation* (GCV) merupakan salah satu metode yang sangat sering digunakan untuk memilih titik knot dan *bandwidth* yang paling optimal. Metode GCV merupakan bentuk modifikasi dari metode CV. Model regresi nonparametrik dengan estimator campuran yaitu fungsi *spline truncated* dan fungsi kernel Gaussian yang terbaik dengan titik knot dan *bandwidth* optimal didapatkan dari nilai GCV minimum (Budiantara, 2019). Fungsi GCV diberikan oleh:

$$GCV(\phi_{opt}, \alpha_{opt}) = \left[ \frac{MSE(\phi, \alpha)}{(n^{-1} \text{trace}\{\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]\})^2} \right] \quad (2.24)$$

Dengan  $\mathbf{I}$  merupakan matriks identitas dan  $MSE(\phi, \alpha)$  sebagai berikut:

$$MSE(\phi, \alpha) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.25)$$

Titik knot dan *bandwidth* paling optimal dalam  $GCV(\phi_{opt}, \alpha_{opt})$  adalah dengan meminimumkan fungsi GCV pada Persamaan (2.25) (Green dan Silverman, 1994).

Secara umum, Wu dan Zhang (2006) menjadikan persamaan GCV sebagai berikut:

$$\begin{aligned} GCV(\phi_{opt}, \alpha_{opt}) &= \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left[ n^{-1} \text{trace} \{ \mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)] \} \right]^2} \\ &= \frac{n^{-1} \left[ \bar{\mathbf{Y}}^T (\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)])^T (\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]) \bar{\mathbf{Y}} \right]}{\left[ n^{-1} \text{trace} \{ \mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)] \} \right]^2} \quad (2.26) \\ &= \frac{n^{-1} \left[ \bar{\mathbf{Y}}^T (\mathbf{I} - [\mathbf{B}(\phi, \alpha)])^T (\mathbf{I} - [\mathbf{B}(\phi, \alpha)]) \bar{\mathbf{Y}} \right]}{\left[ n^{-1} \text{trace} \{ \mathbf{I} - [\mathbf{B}(\phi, \alpha)] \} \right]^2} \end{aligned}$$

### 2.8.3 Metode *Unbiased Risk* (UBR)

Metode *Unbiased Risk* (UBR) adalah salah satu metode yang juga dapat digunakan untuk memilih titik knot dan *bandwidth* yang optimal dalam estimator campuran fungsi *spline truncated* dan fungsi kernel Gaussian dengan menambahkan informasi mengenai  $\sigma^2$ . Nilai UBR minimum yang dihasilkan akan memberikan titik knot dan *bandwidth* yang optimal. Selanjutnya formula dari UBR diberikan pada Persamaan (2.27).

$$\begin{aligned} UBR(\phi_{opt}, \alpha_{opt}) &= n^{-1} \left\{ \begin{aligned} &\left\| (\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]) \bar{\mathbf{Y}} \right\|^2 + \\ &\frac{\hat{\sigma}^2}{n} \text{trace} [\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]]^2 + \\ &\frac{\hat{\sigma}^2}{n} \text{trace} [ [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]^2 ]^2 \end{aligned} \right\} \\ &= n^{-1} \left\{ \begin{aligned} &\left\| (\mathbf{I} - \mathbf{B}(\phi, \alpha)) \bar{\mathbf{Y}} \right\|^2 + \frac{\hat{\sigma}^2}{n} \text{trace} [\mathbf{I} - \mathbf{B}(\phi, \alpha)]^2 + \\ &\frac{\hat{\sigma}^2}{n} \text{trace} [\mathbf{B}(\phi, \alpha)^2]^2 \end{aligned} \right\} \quad (2.27) \end{aligned}$$

dengan

$$\hat{\sigma}^2 = \frac{\|(\mathbf{I} - \mathbf{B}(\phi, \alpha))\vec{Y}\|^2}{\text{trace}[(\mathbf{I} - \mathbf{B}(\phi, \alpha))\vec{Y}]} \quad (2.28)$$

dan  $\mathbf{B}(\phi, \alpha)$  seperti pada Persamaan (2.22). Berdasarkan ke tiga metode pemilihan titik knot yang optimal, yang membedakan fungsi dari turunan CV yaitu GCV dengan UBR adalah adanya taksiran varians (Devi, dkk., 2018).

## 2.7 Kriteria Pemilihan Model Terbaik

Tujuan utama dalam pemodelan statistika adalah mendapatkan model yang terbaik. Dalam kasus ini, maka semakin baik model yang diperoleh melalui analisis regresi, maka model tersebut diharapkan semakin mampu untuk menjelaskan hubungan antara variabel prediktor terhadap variabel responnya. Salah satu kriteria yang sering digunakan dalam pemilihan model terbaik adalah dengan menggunakan koefisien determinasi ( $R^2$ ).

Koefisien determinasi ( $R^2$ ) itu sendiri bermakna sebagai sumbangan pengaruh yang diberikan oleh variabel prediktor terhadap variabel respon, atau dengan kata lain, nilai  $R^2$  dapat digunakan untuk melihat seberapa besar kontribusi pengaruh yang diberikan variabel prediktor secara bersama-sama terhadap variabel respon. Adapun formula untuk mencari nilai  $R^2$  adalah sebagai berikut:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2.29)$$

Model yang baik adalah model yang memiliki nilai  $R^2$  yang tinggi (Damaliana, dkk., 2019).

## 2.8 Kolmogorov-Smirnov Test

Tujuan dilakukan pengujian asumsi ini adalah untuk mengetahui dan memeriksa apakah residual yang didapatkan dari model telah memenuhi asumsi kenormalan atau tidak. Residual dari model regresi harus mengikuti distribusi normal dengan *mean* nol dan varians  $\sigma^2$ . Pada penelitian ini, metode yang

digunakan untuk melakukan uji asumsi kenormalan residual adalah Kolmogorov-Smirnov. Adapun tahapan dalam pengujian kenormalan residual sebagai berikut:

a. Merumuskan hipotesis

$H_0$  : Residual berdistribusi Normal

$H_1$  : Residual tidak berdistribusi Normal

b. Menentukan nilai tingkat signifikansi yang digunakan

c. Statistik uji yang digunakan yaitu statistik uji Kolmogorov-Smirnov

$$D^* = \sup\{|F_n(z) - F_0(z)|, -\infty \leq z \leq \infty\} \quad (2.30)$$

Dengan  $F_n(z)$  merupakan fungsi distribusi empiris (*empirical distribution function*) sedangkan  $F_0(z)$  adalah fungsi distribusi kumulatif (*cumulative distribution function*) dari normal baku.

d. Daerah kritis yaitu menolak  $H_0$  jika  $D^* > D_{(\alpha,n)}$  atau  $p\text{-value} < \alpha$ .

e. Menarik keputusan dan membuat kesimpulan

## 2.9 Variance Inflation Factor

Multikolinearitas adalah suatu kondisi dimana terjadi korelasi antara variabel bebas. Dalam proses pemodelan menggunakan analisis regresi, tidak diperkenankan terdapat masalah atau kasus multikolinearitas. Ukuran yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah *Variance Inflation Factor* (VIF). Nilai VIF lebih besar dari 10 mengidentifikasi adanya masalah multikolinearitas pada model. Adapun formula untuk mencari nilai VIF sebagai berikut:

$$VIF_p = \frac{1}{1 - R_p^2} \quad (2.31)$$

Dengan  $R_p^2$  merupakan koefisien determinasi antara variabel  $x_p$  dengan variabel bebas lainnya, dengan  $p = 1, 2, \dots, q$  (Sriningsih, dkk., 2018).

## 2.10 Tinjauan Kemiskinan

Pada sub-bab bagian tinjauan kemiskinan ini akan diuraikan dan dijabarkan mengenai konsep dan definisi dari kemiskinan berdasarkan pendapat para ahli,



pengukuran kemiskinan dan juga faktor-faktor yang diduga mempengaruhi kemiskinan.

### **2.10.1 Definisi Kemiskinan**

Secara etimologis, “kemiskinan” berasal dari kata “miskin” yang artinya tidak berharta benda dan serba kekurangan. Badan Pusat Statistik mendefinisikan sebagai ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak, selanjutnya kemiskinan merupakan sebuah kondisi yang berada dibawah garis nilai standar kebutuhan minimum, baik untuk makanan dan non-makanan yang disebut garis kemiskinan (*poverty line*) atau disebut juga batas kemiskinan (*poverty threshold*).

Definisi dari kemiskinan itu sendiri saat ini telah mengalami perluasan, seiring dengan semakin kompleksnya faktor penyebab, indikator serta permasalahan lain-nya yang melingkupi. Kemiskinan tidak lagi hanya dianggap sebagai dimensi ekonomi, melainkan telah meluas hingga ke dimensi sosial, kesehatan, pendidikan hingga politik.

Menurut Yacoub (2012) dalam Ferezagia (2018) menjelaskan bahwa kemiskinan merupakan masalah global, karena kemiskinan merupakan masalah yang dihadapi oleh banyak negara, khususnya negara berkembang. Hidup dalam kemiskinan bukan hanya hidup dalam kekurangan uang dan tingkat pendapatan yang rendah, tetapi juga banyak hal lain.

Kartasasmita (1996) menjelaskan penyebab terjadinya kemiskinan akibat dari berbagai faktor, yang terdiri dari:

1. Rendahnya tingkat pendidikan menyebabkan pengembangan diri yang terbatas.
2. Rendahnya tingkat kesehatan, dimana tingkat kesehatan dan gizi yang rendah menyebabkan daya tahan fisik, daya pikir serta prakarsa menjadi rendah. Dengan demikian produktivitas yang dihasilkan menjadi berkurang.
3. Terbatasnya lapangan pekerjaan. Selama lapangan pekerjaan atau kegiatan usaha masih tersedia, maka harapan untuk memutuskan lingkaran kemiskinan masih dapat dilakukan.
4. Kondisi terisolasi, dimana dalam kondisi terpencil atau terisolasi penduduk akan kurang mampu menjalankan roda perekonomiannya.

Tambunan (2001) dalam penelitian Jacobus, dkk., (2018) menjelaskan banyak sekali faktor yang mempengaruhi secara langsung maupun tidak langsung terhadap kemiskinan, mulai dari produktivitas tenaga kerja, upah, distribusi pendapatan, kesempatan kerja, inflasi, pajak dan subsidi, investasi, alokasi serta sumber daya alam, ketersediaan fasilitas umum (seperti pendidikan dasar, kesehatan, transportasi, listrik, air bersih dan lokasi pemukiman), penggunaan teknologi, tingkat dan jenis pendidikan, kondisi fisik dan alam suatu wilayah, dan lain sebagainya. Bentuk kemiskinan multidimensional inilah yang selanjutnya juga diterangkan sebagai faktor-faktor yang memiskinkan (Suryawati, 2005).

### **2.10.2 Konsep dan Pengukuran Kemiskinan**

Pengukuran tingkat kemiskinan di Indonesia pertama kali secara resmi dipublikasikan BPS pada Tahun 1984 yang mencakup data kemiskinan periode 1976-1981. Penduduk miskin adalah penduduk yang berada di bawah suatu batas, yang disebut batas kemiskinan atau garis kemiskinan. Berdasarkan hasil Widyakarya Pangan dan Gizi 1978, seseorang dapat dikatakan hidup sehat apabila telah dapat memenuhi kebutuhan energinya minimal sebesar 2.100 kilokalori perhari. Mengacu pada ukuran tersebut, maka garis kemiskinan untuk makanan adalah nilai rupiah yang harus dikeluarkan seseorang dalam sebulan agar dapat memenuhi kebutuhan energinya sebesar 2.100 kalori perhari. Agar seseorang dapat hidup layak, pemenuhan akan kebutuhan makanan saja tidak akan cukup, oleh karena itu perlu dipenuhi kebutuhan dasar non-makanan, seperti perumahan, pendidikan, kesehatan, pakaian, serta aneka barang dan jasa lainnya. Ringkasnya, garis kemiskinan terdiri atas dua komponen, yaitu garis kemiskinan makanan dan non-makanan (BPS, 2017).

Garis Kemiskinan (GK) merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non-Makanan (GKNM). Penduduk yang memiliki rata-rata pengeluaran perkapita per bulan dibawah GKM akan dikategorikan sebagai penduduk miskin. Paket komoditi kebutuhan dasar makanan diwakili oleh 52 jenis komoditi (padi-padian, umbi-umbian, ikan, daging, telur dan susu, sayuran, kacang-kacangan, buah-buahan, minyak dan lemak, dan lain sebagainya). GKNM merupakan kebutuhan minimum untuk perumahan, sandang,

pendidikan dan kesehatan. Paket komoditi kebutuhan dasar non-makanan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di pedesaan. Formula untuk menghitung GK adalah sebagai berikut:

$$GK = GKM + GKNM \quad (2.32)$$

Persentase penduduk miskin adalah persentase penduduk yang berada dibawah Garis Kemiskinan (GK). Persentase penduduk miskin merupakan besaran angka penduduk yang penghasilannya atau konsumsinya dibawah garis kemiskinan, yaitu kelompok populasi yang tidak mampu membeli satu paket bahan kebutuhan pokok. Formula untuk menghitung persentase penduduk miskin adalah sebagai berikut:

$$P_a = \frac{1}{n} \sum_{i=1}^c \left[ \frac{b - s_i}{b} \right]^a \quad (2.33)$$

dimana

- $b$  : garis kemiskinan
- $s_i$  : rata-rata pengeluaran perkapita sebulan penduduk yang berada dibawah garis kemiskinan  $i = 1, 2, \dots, c$  dan  $s_i < b$
- $P_a$  : persentase penduduk miskin,  $a = 0$
- $n$  : jumlah penduduk
- $c$  : banyaknya penduduk yang berada dibawah garis kemiskinan

(BPS, 2017)

### 2.10.3 Faktor-Faktor yang diduga Mempengaruhi Kemiskinan

Sejumlah variabel dapat dipakai untuk menduga persoalan kemiskinan, sehingga diharapkan dari variabel-variabel tersebut dapat menghasilkan serangkaian strategi, kebijakan dan penanggulangan kemiskinan yang tepat sasaran dan berkesinambungan.

Muda, dkk. (2019) menjelaskan bahwa intervensi untuk memperbaiki kesehatan dari pemerintah juga merupakan suatu alat kebijakan penting untuk mengurangi kemiskinan. Angka Harapan Hidup (AHH), dapat dijadikan suatu indikator dalam mengukur kesehatan suatu individu di suatu wilayah atau daerah. AHH mencerminkan derajat kesehatan suatu masyarakat. AHH adalah rata-rata perkiraan banyak tahun yang dapat ditempuh seseorang selama hidup. AHH

diartikan sebagai umur yang mungkin dicapai seseorang yang lahir pada tahun tertentu. AHH merupakan alat untuk mengevaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk pada umumnya (BPS, 2017).

Cutler dan Katz (1991) melakukan penelitian pengaruh variabel ekonomi makro seperti pengangguran terhadap kemiskinan, hasilnya menunjukkan bahwa pengangguran memberikan pengaruh yang signifikan dan positif terhadap tingkat kemiskinan. Pengangguran adalah seseorang yang sudah digolongkan dalam angkatan kerja yang secara aktif sedang mencari pekerjaan, tetapi mereka belum dapat memperoleh pekerjaan tersebut. Hartanto dan Masjkuri (2017) dalam penelitiannya menjelaskan bahwa pengangguran yang tinggi mempunyai dampak buruk terhadap perekonomian, individu dan masyarakat, seperti tingginya jumlah pengangguran akan menyebabkan masyarakat tidak dapat memaksimalkan kesejahteraan yang mungkin dicapai, produktivitas dan pendapatan masyarakat akan berkurang sehingga akan memicu semakin parahnya tingkat kemiskinan.

Tingkat Pengangguran Terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Formula untuk menghitung TPT adalah sebagai berikut:

$$TPT = \frac{u}{v} \times 100\% \quad (2.34)$$

Dengan  $u$  adalah jumlah pengangguran dan  $v$  merupakan jumlah angkatan kerja.

Ustama (2009) menjelaskan bahwa pendidikan berperan penting dalam menanggulangi serta mengurangi kemiskinan dalam jangka panjang, baik secara tidak langsung melalui perbaikan produktivitas, maupun secara langsung melalui pelatihan golongan miskin dengan keterampilan yang dibutuhkan untuk meningkatkan produktivitas mereka dan pada waktunya nanti akan meningkatkan pendapatan mereka. Rata-rata lama sekolah mengindikasikan makin tingginya pendidikan yang dicapai oleh masyarakat di suatu daerah.

Menurut BPS, rata-rata lama sekolah adalah rata-rata jumlah tahun yang dihabiskan oleh penduduk yang berusia 15 tahun keatas untuk menempuh semua jenis pendidikan formal yang pernah dijalani. Untuk mereka yang tamat SD diperhitungkan lama sekolah selama 6 tahun, tamat SMP diperhitungkan lama sekolah selama 9 tahun, tamat SMA diperhitungkan lama sekolah selama 12 tahun

tanpa memperhitungkan apakah pernah tinggal kelas atau tidak. Formula untuk menghitung rata-rata lama sekolah adalah sebagai berikut:

$$RLS = \frac{1}{P_{15+}} \sum_{i=1}^{P_{15+}} (l_i) \quad (2.35)$$

dimana  $P_{15+}$  adalah jumlah penduduk yang berusia 15 tahun ke atas dan  $l_i$  adalah lama sekolah penduduk ke- $i$ .

Pertumbuhan ekonomi adalah salah satu indikator yang digunakan untuk mengukur prestasi ekonomi suatu negara. Ginting dan Dewi (2013) menjelaskan bahwa pertumbuhan ekonomi merupakan syarat keharusan bagi pengurangan kemiskinan. Syahputra (2017) menjelaskan bahwa pertumbuhan ekonomi adalah proses kenaikan *output* per kapita dalam jangka panjang. Pertumbuhan ekonomi akan dapat diwujudkan apabila pertumbuhan PDRB riil melebihi laju pertumbuhan penduduk. Untuk kepentingan analisis ekonomi regional, biasanya digunakan pertumbuhan Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan.

Standar kesehatan mencerminkan pola kehidupan keseharian dari masyarakat. Kemiskinan akan menjadikan masyarakat tidak dapat memenuhi kualitas standar dari kehidupan sesuai dengan SDGs. Derajat kesehatan yang tinggi harus didukung dengan fasilitas kesehatan dan sanitasi yang baik. Adhi (2009) menjelaskan bahwa pembangunan sanitasi yang layak dapat meningkatkan kesehatan masyarakat dan lingkungan. Mungkasa (2004) dalam penelitian Rizki dan Saleh (2007) menjelaskan bahwa peningkatan kualitas dan ketersediaan sanitasi yang baik dapat meningkatkan kesejahteraan penduduk yang berarti mengurangi tingkat kemiskinan.

Formula untuk menghitung persentase rumah tangga yang memiliki akses terhadap layanan sanitasi layak adalah sebagai berikut:

$$RTSL = \frac{\text{Jumlah RT yang memiliki akses sanitasi layak}}{\text{Jumlah RT}} \times 100\% \quad (2.36)$$

Selanjutnya Putra dan Rianto (2016) dalam penelitian menjelaskan bahwa ketersediaan air minum layak dan bersih masih menjadi permasalahan umum di dunia, sehingga masih menjadi target pembangunan dunia yang tertuang dalam SDGs. Indonesia khususnya masih menghadapi permasalahan ketersediaan air minum yang layak dan bersih, drainase serta sanitasi. Chong, dkk. (2016)

menjelaskan rendahnya kualitas dan kuantitas sarana dan prasarana akses air minum layak dan bersih akan mempengaruhi tingkat kemiskinan di suatu daerah. Berdasarkan hasil penelitiannya menunjukkan bahwa rumah tangga yang tidak memiliki akses terhadap air minum layak dan bersih akan memiliki kecenderungan yang lebih tinggi menjadi keluarga miskin.

Formula untuk menghitung persentase rumah tangga yang memiliki akses air minum layak adalah sebagai berikut:

$$RTAML = \frac{\text{Jumlah RT yang memiliki akses air minum layak}}{\text{Jumlah RT}} \times 100\% \quad (2.37)$$

Didu dan Fauzi (2016) mengemukakan bahwa terdapat pengaruh langsung antara jumlah penduduk terhadap tingkat kesejahteraan masyarakat. Pertumbuhan jumlah penduduk yang pesat khususnya di negara berkembang menyebabkan penurunan tingkat kesejahteraan masyarakat. Selanjutnya Sartika, dkk. (2016) menjelaskan bahwa faktor dasar yang menyebabkan kemiskinan adalah kepadatan penduduk. Kepadatan penduduk menggambarkan jumlah penduduk yang terpusat dalam suatu wilayah tertentu. Nyompa, dkk. (2019) dalam penelitiannya menjelaskan bahwa kepadatan penduduk yang berlebih dianggap sebagai beban pembangunan, memperkecil pendapatan perkapita serta akan menimbulkan masalah ketenagakerjaan.

## BAB 3

### METODOLOGI PENELITIAN

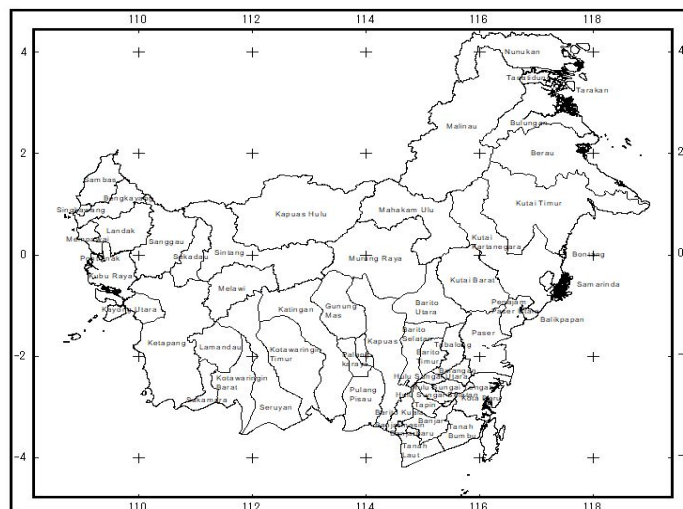
#### 3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder Tahun 2019 yang dipublikasikan oleh Badan Pusat Statistik (BPS). Unit observasi yang digunakan adalah Kabupaten/Kota yang ada di Pulau Kalimantan yang terdiri atas 5 Provinsi saat ini.

**Tabel 3.1** Jumlah Kabupaten/Kota di Pulau Kalimantan berdasarkan Provinsi

	Provinsi	Jumlah Kabupaten/Kota
Pulau Kalimantan	Kalimantan Barat	14
	Kalimantan Selatan	13
	Kalimantan Utara	5
	Kalimantan Tengah	14
	Kalimantan Timur	10
<b>Total Kabupaten/Kota</b>		<b>56</b>

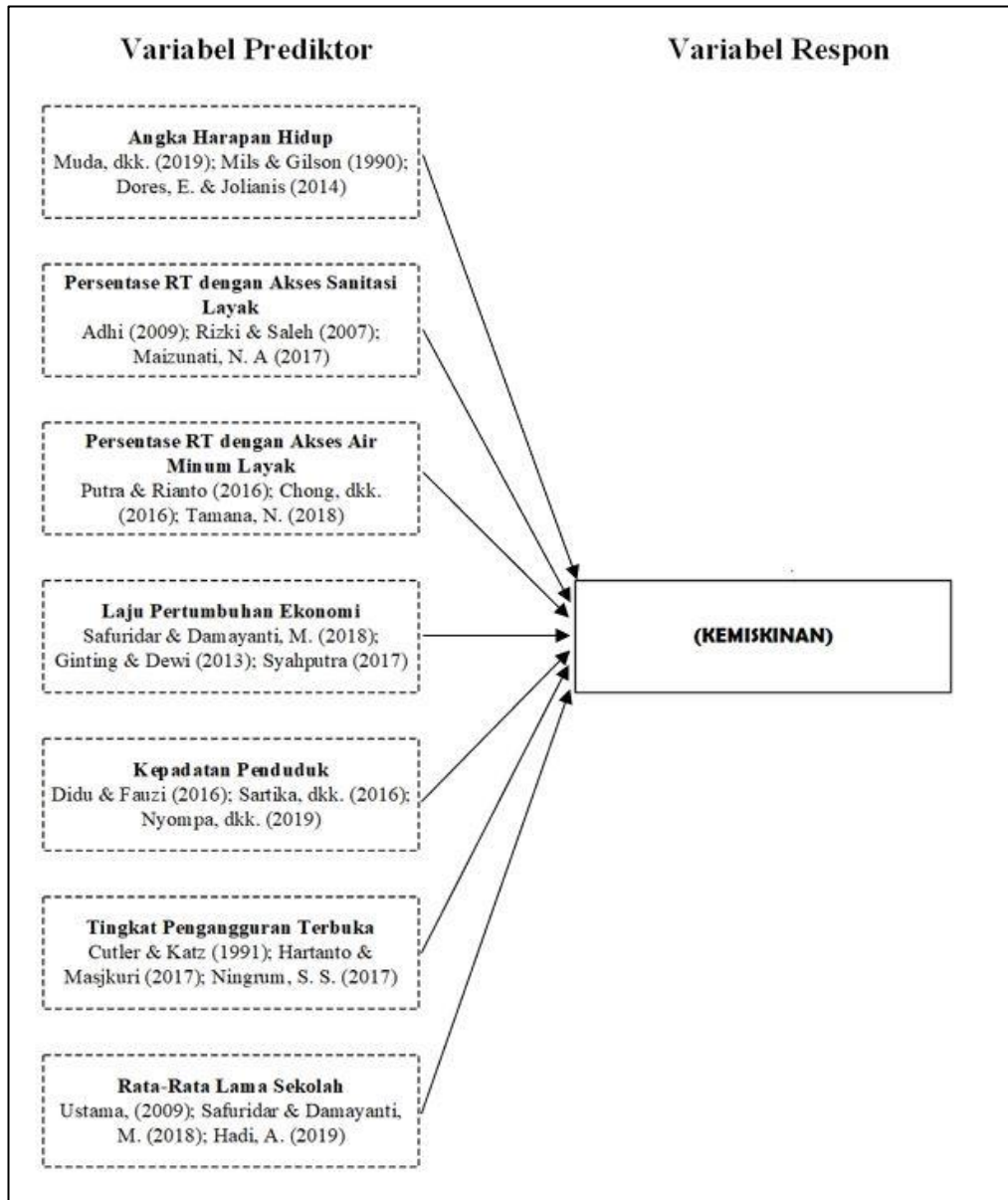
Pada penelitian ini, Kota Bontang, Banjarmasin dan Pontianak tidak digunakan dalam observasi penelitian karena merupakan data pencilan, sehingga banyaknya Kabupaten/Kota yang dijadikan observasi pengamatan pada penelitian ini tersisa 53 Kabupaten/Kota. Letak atau posisi dari setiap Kabupaten/Kota yang ada di Pulau Kalimantan dapat dilihat pada Gambar 3.1.



**Gambar 3.1** Peta pulau Kalimantan

### 3.2 Kerangka Konseptual Penelitian

Berdasarkan penjelasan mengenai tinjauan kemiskinan pada Bab 2, terdapat beberapa faktor yang diduga mempengaruhi kemiskinan. Lebih jelasnya, kerangka konseptual penelitian di visualisasikan seperti pada Gambar 3.2.



Gambar 3.2 Variabel yang diduga mempengaruhi kemiskinan

### 3.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon dan variabel prediktor. Dalam penelitian ini digunakan satu variabel respon dan tujuh



variabel prediktor. Variabel-variabel yang digunakan dalam penelitian ini diberikan dalam Tabel 3.2 sebagai berikut:

**Tabel 3.2** Variabel Penelitian

Variabel	Simbol Variabel	Deskripsi	Satuan
Respon	$y$	Persentase Penduduk Miskin	Persen
	$x_1$	Angka Harapan Hidup	Tahun
Prediktor	$x_2$	Tingkat Pengangguran Terbuka	Persen
	$x_3$	Rata-Rata Lama Sekolah	Tahun
	$x_4$	Laju Pertumbuhan Ekonomi	Persen
	$x_5$	Persentase RT dengan Akses Sanitasi Layak	Persen
	$x_6$	Persentase RT dengan Akses Air Minum Layak	Persen
	$t_1$	Kepadatan Penduduk	Jiwa/Km <sup>2</sup>

Definisi operasional dari variabel penelitian di atas adalah sebagai berikut:

1. **Persentase Penduduk Miskin (Y)**

Persentase penduduk miskin adalah persentase penduduk yang berada dibawah garis kemiskinan. Garis kemiskinan merupakan penjumlahan dari garis kemiskinan makanan (GKM) dan garis kemiskinan non-makanan (GKNM). Adapun formula untuk menghitung GK ada pada Persamaan (2.32), sedangkan formula untuk menghitung persentase penduduk miskin pada Persamaan (2.33).

2. **Angka Harapan Hidup (X<sub>1</sub>)**

Angka harapan hidup (AHH) adalah rata-rata perkiraan banyak tahun yang dapat ditempuh seseorang selama hidup. AHH diartikan sebagai umur yang mungkin dicapai seseorang yang lahir pada tahun tertentu.

3. **Tingkat Pengangguran Terbuka (X<sub>2</sub>)**

Tingkat pengangguran terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Penduduk yang termasuk angkatan kerja adalah penduduk usia kerja (15 tahun dan lebih) yang bekerja atau punya pekerjaan namun sementara tidak bekerja atau pengangguran. Adapun formula untuk menghitung TPT ada pada Persamaan (2.34).

4. **Rata-rata Lama Sekolah ( $X_3$ )**

Rata-rata lama sekolah adalah rata-rata jumlah tahun yang dihabiskan oleh penduduk yang berusia 15 tahun ke atas untuk menempuh semua jenis pendidikan formal yang pernah dijalani. Rata-rata lama sekolah mengindikasikan makin tingginya pendidikan yang dicapai oleh masyarakat di suatu daerah. Adapun formula untuk menghitung rata-rata lama sekolah ada pada Persamaan (2.35).

5. **Laju Pertumbuhan Ekonomi ( $X_4$ )**

Laju pertumbuhan ekonomi diukur dengan menggunakan laju pertumbuhan PDRB atas dasar harga konstan. PDRB atas dasar harga konstan menunjukkan nilai tambah barang dan jasa tersebut yang dihitung menggunakan harga yang berlaku pada satu tahun tertentu sebagai tahun dasar. Laju pertumbuhan ekonomi yang digunakan dalam penelitian ini adalah PDRB atas dasar harga konstan tahun 2010. Laju pertumbuhan ekonomi adalah pertambahan PDRB suatu kabupaten yang berlaku dari tahun ke tahun yang menunjukkan tingkat perkembangan riil dari agregat pendapatan untuk masing-masing tahun dibandingkan dengan tahun sebelumnya.

6. **Persentase Rumah Tangga dengan Akses Sanitasi Layak ( $X_5$ )**

Merupakan perbandingan antara jumlah rumah tangga yang memiliki akses sanitasi layak terhadap jumlah rumah tangga, yang dinyatakan dalam persen. Ketentuan sanitasi layak yang memenuhi syarat kesehatan yaitu digunakan oleh rumah tangga sendiri atau bersama dengan orang lain tertentu, dilengkapi dengan kloset jenis leher angsa, serta tempat pembuangan akhir tinja. Adapun formula untuk menghitung persentase rumah tangga yang memiliki akses terhadap sanitasi layak ada pada Persamaan (2.36).

7. **Persentase Rumah Tangga dengan Akses Air Minum Layak ( $X_6$ )**

Air minum yang berkualitas (layak) adalah air minum yang terlindung meliputi air ledeng (keran), keran umum, *hydrant* umum, terminal air, penampungan air hujan (PAH) atau mata air dan sumur terlindung, sumur bor atau sumur pompa, yang jaraknya minimal 10 m dari pembuangan kotoran, penampungan limbah dan pembuangan sampah. Persentase rumah tangga

dengan akses terhadap air minum layak adalah perbandingan antara rumah tangga dengan akses terhadap sumber air minum berkualitas (layak) dengan rumah tangga seluruhnya, yang kemudian dinyatakan dalam persentase.

#### 8. **Kepadatan Penduduk ( $T_1$ )**

Kepadatan penduduk adalah suatu keadaan yang dikatakan semakin padat apabila jumlah manusia pada suatu batas ruang tertentu semakin banyak dibandingkan dengan luas ruangnya. Secara umum, kepadatan penduduk menunjukkan banyaknya jumlah penduduk untuk setiap kilometer persegi luas wilayah. Selanjutnya kepadatan penduduk dapat dijadikan indikator dari pada tekanan penduduk di suatu daerah.

### 3.4 **Struktur Data Penelitian**

Struktur data dari penelitian ini disajikan pada Tabel 3.3 sebagai berikut:

**Tabel 3.3** Struktur Data Penelitian

Variabel Respon ( $y_i$ )	Variabel Komponen <i>Spline Truncated</i> ( $x_{qi}$ )				Variabel Komponen Kernel ( $t_i$ )
	$x_{1i}$	$x_{2i}$	...	$x_{6i}$	$t_{1i}$
$y_1$	$x_{1.1}$	$x_{2.1}$	...	$x_{6.1}$	$t_{1.1}$
$y_2$	$x_{1.2}$	$x_{2.2}$	...	$x_{6.2}$	$t_{1.2}$
$y_3$	$x_{1.3}$	$x_{2.3}$	...	$x_{6.3}$	$t_{1.3}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$y_{53}$	$x_{1.53}$	$x_{2.53}$	...	$x_{6.53}$	$t_{1.53}$

### 3.5 **Tahapan Penelitian**

Untuk menyelesaikan tujuan penelitian, maka perlu disusun langkah-langkah atau tahapan penelitian. Adapun langkah-langkah yang digunakan pada penelitian ini adalah:

1. **Tujuan ke-1** adalah mendapatkan estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian.

Untuk menyelesaikan tujuan ini, maka dilakukan langkah-langkah sebagai berikut:

- a. Mendefinisikan model regresi nonparametrik *spline truncated* dan kernel Gaussian. Diberikan data berpasangan dengan variabel prediktor sebanyak

6 untuk komponen *spline truncated* dan 1 untuk komponen kernel sedemikian sehingga dapat ditulis  $(x_{1i}, x_{2i}, \dots, x_{6i}, t_i, y_i)$ . Asumsikan bahwa pola hubungan antara variabel prediktor  $(x_{1i}, x_{2i}, \dots, x_{6i}, t_i)$  dan variabel respon  $y_i$  mengikuti model regresi nonparametrik yaitu  $y_i = f(x_{1i}, x_{2i}, \dots, x_{6i}) + h(t_i) + \varepsilon_i$  dengan  $i = 1, 2, \dots, n$ .

Kurva regresi dari komponen *spline truncated* yaitu  $f(x_{1i}, x_{2i}, \dots, x_{6i})$  diasumsikan bersifat aditif, sehingga didapatkan

$$\begin{aligned} f(x_{1i}, x_{2i}, \dots, x_{6i}) &= f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_6(x_{6i}) \\ &= \sum_{p=1}^6 f_p(x_{pi}), i = 1, 2, \dots, n \end{aligned}$$

Selanjutnya model regresi nonparametrik *spline truncated* dan kernel Gaussian dapat dituliskan  $y_i = \sum_{p=1}^6 f_p(x_{pi}) + h(t_i) + \varepsilon_i$ .

- b. Menghampiri kurva regresi nonparametrik  $f_p(x_{pi})$  dengan  $p = 1, 2, \dots, 6$  menggunakan estimator fungsi *spline truncated*.
- c. Menghampiri kurva regresi nonparametrik  $h(t_i)$  dengan  $i = 1, 2, \dots, n$  menggunakan estimator fungsi kernel Gaussian.
- d. Menyajikan model estimasi campuran regresi nonparametrik *spline truncated* dan kernel Gaussian ke dalam bentuk matriks:

Diketahui sebelumnya jika  $y_i = \mu(x_i, t_i) + \varepsilon_i, i = 1, 2, \dots, n$  dengan

$$\mu_{\alpha, \phi}(x_i, t_i) = \sum_{p=1}^q f_p(x_{pi}) + h_{\alpha}(t_i).$$

Sehingga apabila dilakukan substitusi akan didapatkan

$$y_i = \left[ \sum_{p=1}^q f_p(x_{pi}) + h_{\alpha}(t_i) \right] + \varepsilon_i, i = 1, 2, \dots, n.$$

Kemudian disajikan dalam bentuk matriks menjadi  $\vec{Y} = \mathbf{X}(\phi)\vec{\beta} + \mathbf{D}(\alpha)\vec{Y} + \vec{\varepsilon}$ .

Dengan  $\vec{Y}$  merupakan vektor variabel respon,  $\mathbf{X}$  merupakan matriks komponen *spline truncated*,  $\mathbf{D}$  merupakan matriks komponen *kernel* dan  $\vec{\varepsilon}$  merupakan vektor *error* random.

- e. Mendapatkan estimasi campuran regresi nonparametrik *spline truncated* dan kernel Gaussian menggunakan metode *Ordinary Least Squares* (OLS):

$$\begin{aligned}
Q(\vec{\beta} | \phi, \alpha) &= \sum_{i=1}^n \tilde{\varepsilon}_i^2 \\
&= \tilde{\varepsilon}^T \tilde{\varepsilon} \\
&= \left( [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} - \mathbf{X}(\phi) \vec{\beta} \right)^T \left( [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} - \mathbf{X}(\phi) \vec{\beta} \right) \\
&= \left\| [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} \right\|^2 - 2 \vec{\beta}^T \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi) \vec{\beta}
\end{aligned}$$

- f. Menyelesaikan optimasi dari  $Q(\vec{\beta} | \phi, \alpha)$  sehingga perlu dilakukan derivatif parsial  $Q(\vec{\beta} | \phi, \alpha)$  terhadap  $\vec{\beta}$ .

$$\frac{\partial Q(\vec{\beta} | \phi, \alpha)}{\partial \vec{\beta}} = \frac{\partial \left( \left\| [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} \right\|^2 - 2 \vec{\beta}^T \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi) \vec{\beta} \right)}{\partial \vec{\beta}}$$

- g. Kemudian menyamakan hasil derivatif parsial  $Q(\vec{\beta} | \phi, \alpha)$  dengan 0, sehingga

$$\frac{\partial Q(\vec{\beta} | \phi, \alpha)}{\partial \vec{\beta}} = 0$$

dengan

$$Q(\vec{\beta} | \phi, \alpha) = \left\| [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} \right\|^2 - 2 \vec{\beta}^T \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi) \vec{\beta} .$$

- h. Mendapatkan estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian.

$$\hat{\mu}_{\alpha, \phi}(x_i, t_i) = \sum_{p=1}^q \hat{f}_p(x_{pi}) + \hat{h}_\alpha(t_i)$$

2. **Tujuan ke-2** adalah melakukan studi simulasi pada model estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian untuk membandingkan kinerja metode CV, GCV dan UBR dalam menduga titik knot dan *bandwidth* yang optimal.

Untuk menyelesaikan tujuan ini, maka dilakukan langkah-langkah sebagai berikut:

- a. Mendefinisikan model regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian.

$$y_i = \mu(x_i, t_i) + \varepsilon_i, \text{ dengan } \mu(x_i, t_i) = f(x_i) + h(t_i)$$

Selanjutnya  $f(x_i)$  dan  $h(t_i)$  merupakan fungsi-fungsi yang *smooth*.

- b.  $f(x_i)$  adalah fungsi *smooth* yang di definisikan sebagai komponen *spline truncated*

$$\text{dengan } f(x_i) = \frac{\sin(\pi x_i^2)^4}{\sin(x_i^2)}.$$

Variabel  $x_i$  yang di tetapkan untuk komponen *spline truncated* dibangkitkan secara independen dari distribusi Uniform  $x_i \sim U(0,1)$ .

- c.  $h(t_i)$  adalah fungsi yang di definisikan sebagai komponen kernel Gaussian dengan  $h(t_i) = t_i^2 \sin t_i$ .

Variabel  $t_i$  yang di tetapkan untuk komponen kernel Gaussian dibangkitkan secara independen dari distribusi Uniform  $t_i \sim U(0,1)$ .

- d. Kemudian dibangkitkan *error* atau sisaan yaitu  $\varepsilon_i$ , dimana  $i = 1, 2, \dots, n$  yang mengikuti distribusi Normal  $(0, \sigma^2)$ , sehingga  $\varepsilon_i \sim N(0, \sigma^2)$ .

- e. Maka akan didapatkan variabel respon  $y_i$  dari model regresi nonparametrik estimator campuran sebagai berikut:

$$y_i = f(x_i) + h(t_i) + \varepsilon_i \rightarrow \left[ \frac{\sin(\pi x_i^2)^4}{\sin(x_i^2)} \right] + [t_i^2 \sin t_i] + \varepsilon_i$$

- f. Membuat variasi ukuran sampel dengan  $n$  yang akan dicobakan adalah 25, 50, 100, 200 dan varians *error*  $\sigma^2$  yang akan dicobakan adalah 0,05; 0,5 dan 1.

- g. Melakukan replikasi untuk setiap data bangkitan sebanyak 20 kali.

- h. Membuat diagram pencar antara variabel respon  $y_i$  dengan masing-masing variabel prediktor  $(x_i, t_i)$ .

- i. Memodelkan data bangkitan pada langkah (a-h) dengan model regresi nonparametrik estimator campuran fungsi *spline truncated* dan Kernel Gaussian.

Secara umum, model regresi nonparametrik dengan estimator campuran *spline truncated* dan kernel Gaussian dapat dituliskan:

$$\begin{aligned}\hat{\mu}_{\phi,\alpha}(x_i, t_i) &= \hat{f}_{\phi}(x_i) + \hat{h}_{\alpha}(t_i) \\ &= [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]\bar{Y} \\ &= \mathbf{B}(\phi, \alpha)\bar{Y}\end{aligned}$$

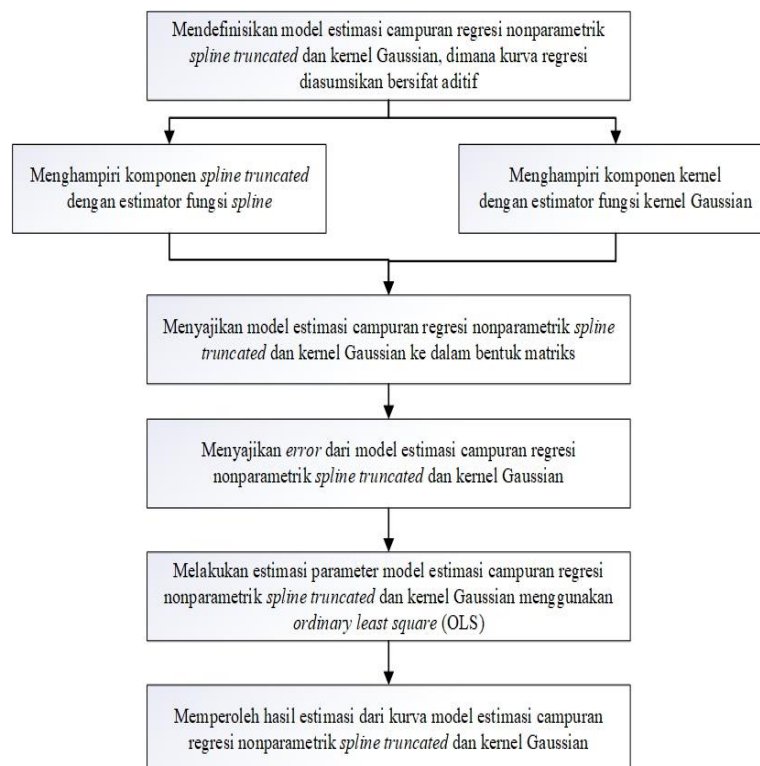
- j. Banyaknya titik knot yang akan dicobakan hanya 1 knot saja, berlaku untuk variabel yang di atur sebagai komponen *spline truncated*.
  - k. Penentuan titik knot dan *bandwidth* yang optimal pada penelitian ini menggunakan metode CV, GCV dan UBR.
  - l. Menghitung nilai koefisien determinasi ( $R^2$ ) untuk masing-masing proses pemodelan yang dilakukan.
  - m. Menghitung rata-rata nilai CV, GCV, UBR dan koefisien determinasi ( $R^2$ ) untuk setiap kombinasi variasi ukuran sampel dan varians error.
  - n. Membuat analisis berdasarkan *output* yang dihasilkan.
3. **Tujuan ke-3** adalah membandingkan metode pemilihan titik knot dan *bandwidth* yang optimal dengan metode CV, GCV, dan UBR dalam aplikasinya menggunakan data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019.

Untuk menyelesaikan tujuan ini, maka dilakukan langkah-langkah sebagai berikut:

- a. Membuat diagram pencar (*scatter plot*) antara variabel respon ( $y_i$ ) dengan masing-masing variabel prediktor ( $x_{1i}, x_{2i}, \dots, x_{6i}, t_i$ ).
- b. Mendeteksi apakah ada hubungan atau korelasi yang kuat antar variabel prediktor, menggunakan nilai *Variance Inflation Factor* (VIF).
- c. Memodelkan data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019 dengan menggunakan estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian.
- d. Mendapatkan optimasi CV berdasarkan Persamaan (2.23).
- e. Menetapkan model terbaik dari nilai CV terkecil dan kemudian menghitung nilai koefisien determinasi ( $R^2$ ).
- f. Mendapatkan optimasi GCV berdasarkan Persamaan (2.24).

- g. Menetapkan model terbaik dari nilai GCV terkecil dan kemudian menghitung nilai koefisien determinasi ( $R^2$ ).
- h. Mendapatkan optimasi UBR berdasarkan Persamaan (2.27) dan (2.28).
- i. Menetapkan model terbaik dari nilai UBR terkecil dan kemudian menghitung nilai koefisien determinasi ( $R^2$ ).
- j. Membandingkan kebaikan model estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian berdasarkan pemilihan titik knot dan *bandwidth* yang optimal menggunakan metode CV, GCV dan UBR dengan kriteria nilai koefisien determinasi ( $R^2$ ).
- k. Melakukan uji asumsi residual yaitu pemeriksaan apakah residual dari model estimator campuran regresi nonparametrik yang terbaik berdasarkan setiap metode pemilihan titik knot dan *bandwidth* optimal mengikuti distribusi normal atau tidak. Statistik uji yang digunakan adalah Kolmogorov-Smirnov *Test*.
- l. Membuat interpretasi model dan selanjutnya membuat kesimpulan.

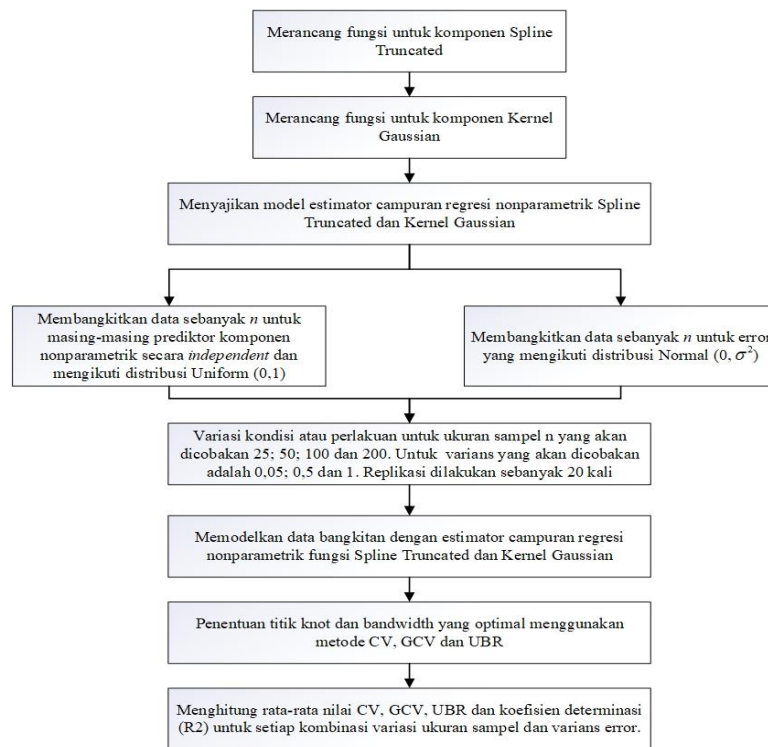
Selanjutnya untuk tujuan 1 dapat di visualisasikan seperti pada Gambar 3.3:



**Gambar 3.3** Langkah analisis untuk tujuan pertama

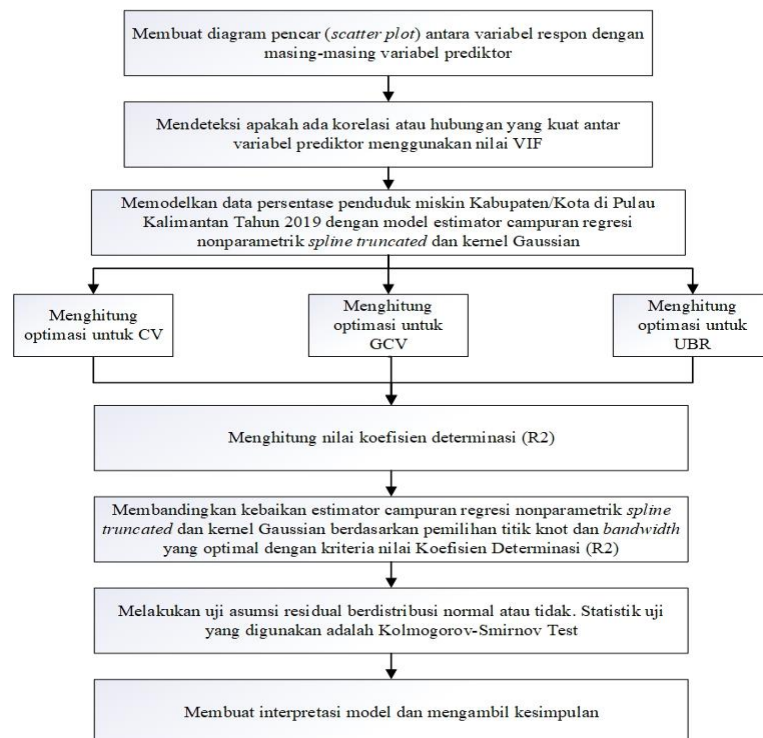


Kemudian untuk tujuan 2 dapat di visualisasikan seperti pada Gambar 3.4:



**Gambar 3.4** Langkah analisis untuk tujuan kedua

Selanjutnya untuk tujuan 3 dapat di visualisasikan seperti pada Gambar 3.5:



**Gambar 3.5** Langkah analisis untuk tujuan ketiga

*(Halaman ini sengaja dikosongkan)*

## BAB 4

### HASIL DAN PEMBAHASAN

Bab ini akan menyajikan hasil dan pembahasan untuk menjawab Tujuan 1, 2 dan 3 yang dituliskan pada Bab 1. Tujuan 1 di fokuskan pada bentuk estimasi kurva regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel. *Spline truncated* yang digunakan adalah *spline truncated* multivariabel dengan fungsi linier, sedangkan untuk kernel akan menggunakan fungsi Gaussian. Tujuan 2 di fokuskan pada kajian simulasi dari metode CV, GCV dan UBR yang diterapkan pada data simulasi. Selanjutnya, tujuan 3 di fokuskan pada aplikasi data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019.

#### 4.1 Model Regresi Nonparametrik Estimator Campuran *Spline Truncated* dan Kernel Gaussian

Berdasarkan tujuan penelitian pertama, maka akan dilakukan kajian mengenai regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian.

##### 4.1.1 Estimator Kurva Regresi untuk Fungsi *Spline Truncated* Multivariabel

Kajian mengenai regresi nonparametrik *spline truncated* multivariabel pada data berpasangan  $(x_{1i}, x_{2i}, \dots, x_{qi}, y_i)$  yang diasumsikan hubungan antara variabel prediktor dan variabel respon mengikuti model regresi nonparametrik.

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{qi}) + \varepsilon_i, i = 1, 2, \dots, n \quad (4.1)$$

Bentuk kurva regresi dari  $f(x_{1i}, x_{2i}, \dots, x_{qi})$  tidak diketahui dan hanya di asumsikan bahwa kurva tersebut *smooth* dalam arti kontinu dan *differensiabel*. *Error random*  $\varepsilon_i$  berdistribusi normal dengan rata-rata nol dan  $E(\varepsilon_i^2) = \sigma^2$ . Selanjutnya kurva regresi  $f(x_{1i}, x_{2i}, \dots, x_{qi})$  diasumsikan bersifat aditif, sedemikian sehingga dapat ditulis dalam bentuk:

$$\begin{aligned} f(x_{1i}, x_{2i}, \dots, x_{qi}) &= f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_q(x_{qi}) \\ &= \sum_{p=1}^q f_p(x_{pi}) \end{aligned} \quad (4.2)$$

Kemudian bentuk pola hubungan respon ( $y_i$ ) dengan masing-masing variabel prediktor ( $x_i$ ) diasumsikan berubah-ubah pada sub-sub interval tertentu. Secara teori  $\sum_{p=1}^q f_p(x_{pi})$  merupakan komponen variabel prediktor yang didekati kurva regresi *spline truncated* derajat 1 (linear) dengan jumlah variabel prediktor sebanyak  $q$  variabel prediktor. Kurva regresi  $f_p(x_{pi})$  diasumsikan termuat dalam suatu ruang *spline* derajat  $m$  dengan  $\phi$  merupakan titik-titik knot yaitu  $K_1, K_2, \dots, K_r$ .

Komponen kurva regresi *spline truncated* multivariabel derajat 1 (linear) untuk masing-masing variabel prediktor dapat dituliskan sebagai berikut:

$$\begin{aligned} f_1(x_1) &= \delta_{01} + \delta_{11}x_1 + \phi_{21}(x_1 - K_{11})_+ + \phi_{31}(x_1 - K_{21})_+ + \dots + \phi_{r1}(x_1 - K_{r1})_+ \\ f_2(x_2) &= \delta_{02} + \delta_{12}x_2 + \phi_{22}(x_2 - K_{12})_+ + \phi_{32}(x_2 - K_{22})_+ + \dots + \phi_{r2}(x_2 - K_{r2})_+ \\ &\vdots \\ f_q(x_q) &= \delta_{0q} + \delta_{1q}x_q + \phi_{2q}(x_q - K_{1q})_+ + \phi_{3q}(x_q - K_{2q})_+ + \dots + \phi_{rq}(x_q - K_{rq})_+ \end{aligned}$$

Maka bisa ditulis secara umum sebagai berikut:

$$f_p(x_p) = \delta_{0p} + \sum_{j=1}^m \sum_{p=1}^q \delta_{jp} x_p^j + \sum_{k=1}^r \sum_{p=1}^q \phi_{(m+k)p} (x_p - K_{kp})_+^m \quad (4.3)$$

Kemudian model regresi *spline truncated* multivariabel derajat 1 (linear) dapat dijabarkan kembali menjadi:

$$\begin{aligned} \sum_{p=1}^q f_p(x_{pi}) &= f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_q(x_{qi}) \\ &= [\delta_{01} + \delta_{11}x_{1i} + \phi_{21}(x_{1i} - K_{11})_+ + \phi_{31}(x_{1i} - K_{21})_+ + \dots + \phi_{r1}(x_{1i} - K_{r1})_+] + \\ &\quad [\delta_{02} + \delta_{12}x_{2i} + \phi_{22}(x_{2i} - K_{12})_+ + \phi_{32}(x_{2i} - K_{22})_+ + \dots + \phi_{r2}(x_{2i} - K_{r2})_+] + \dots + \\ &\quad [\delta_{0q} + \delta_{1q}x_{qi} + \phi_{2q}(x_{qi} - K_{1q})_+ + \phi_{3q}(x_{qi} - K_{2q})_+ + \dots + \phi_{rq}(x_{qi} - K_{rq})_+] \end{aligned}$$

Sedemikian sehingga:

$$\begin{aligned}
f_{pi} &= \sum_{p=1}^q f_p(x_{pi}) \\
&= f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_q(x_{qi}) \\
&= \left\{ \begin{aligned} &\delta_{01} + \delta_{02} + \dots + \delta_{0q} + \delta_{11}x_{1i} + \phi_{21}(x_{1i} - K_{11})_+ + \dots + \phi_{r1}(x_{1i} - K_{r1})_+ + \\ &\delta_{12}x_{2i} + \phi_{22}(x_{2i} - K_{12})_+ + \dots + \phi_{r2}(x_{2i} - K_{r2})_+ + \dots + \\ &\delta_{1q}x_{qi} + \phi_{2q}(x_{qi} - K_{1q})_+ + \dots + \phi_{rq}(x_{qi} - K_{rq})_+ \end{aligned} \right\} \\
&= \left\{ \begin{aligned} &\delta_0^* + \delta_{11}x_{1i} + \phi_{21}(x_{1i} - K_{11})_+ + \dots + \phi_{r1}(x_{1i} - K_{r1})_+ + \\ &\delta_{12}x_{2i} + \phi_{22}(x_{2i} - K_{12})_+ + \dots + \phi_{r2}(x_{2i} - K_{r2})_+ + \dots + \\ &\delta_{1q}x_{qi} + \phi_{2q}(x_{qi} - K_{1q})_+ + \dots + \phi_{rq}(x_{qi} - K_{rq})_+ \end{aligned} \right\} \\
&= \left\{ \begin{aligned} &\delta_0^* + \delta_{11}x_{1i} + \delta_{12}x_{2i} + \dots + \delta_{1q}x_{qi} + \phi_{21}(x_{1i} - K_{11})_+ + \dots + \phi_{r1}(x_{1i} - K_{r1})_+ + \\ &\phi_{22}(x_{2i} - K_{12})_+ + \dots + \phi_{r2}(x_{2i} - K_{r2})_+ + \dots + \phi_{2q}(x_{qi} - K_{1q})_+ + \dots + \phi_{rq}(x_{qi} - K_{rq})_+ \end{aligned} \right\}
\end{aligned}$$

dimana  $\delta_0^* = \delta_{01} + \delta_{02} + \dots + \delta_{0q}$

Misalkan untuk  $i = 1$ , maka:

$$\begin{aligned}
f_{p1} &= \delta_0^* + \delta_{11}x_{11} + \delta_{12}x_{21} + \dots + \delta_{1q}x_{q1} + \phi_{21}(x_{11} - K_{11})_+ + \dots + \phi_{r1}(x_{11} - K_{r1})_+ + \\
&\quad \phi_{22}(x_{21} - K_{12})_+ + \dots + \phi_{r2}(x_{21} - K_{r2})_+ + \dots + \phi_{2q}(x_{q1} - K_{1q})_+ + \dots + \phi_{rq}(x_{q1} - K_{rq})_+
\end{aligned}$$

untuk  $i = 2$ , maka:

$$\begin{aligned}
f_{p2} &= \delta_0^* + \delta_{11}x_{12} + \delta_{12}x_{22} + \dots + \delta_{1q}x_{q2} + \phi_{21}(x_{12} - K_{11})_+ + \dots + \phi_{r1}(x_{12} - K_{r1})_+ + \\
&= \phi_{22}(x_{22} - K_{12})_+ + \dots + \phi_{r2}(x_{22} - K_{r2})_+ + \dots + \phi_{2q}(x_{q2} - K_{1q})_+ + \dots + \phi_{rq}(x_{q2} - K_{rq})_+
\end{aligned}$$

dan seterusnya hingga  $i = n$ , maka:

$$\begin{aligned}
f_{pn} &= \delta_0^* + \delta_{11}x_{1n} + \delta_{12}x_{2n} + \dots + \delta_{1q}x_{qn} + \phi_{21}(x_{1n} - K_{11})_+ + \dots + \phi_{r1}(x_{1n} - K_{r1})_+ + \\
&= \phi_{22}(x_{2n} - K_{12})_+ + \dots + \phi_{r2}(x_{2n} - K_{r2})_+ + \dots + \phi_{2q}(x_{qn} - K_{1q})_+ + \dots + \phi_{rq}(x_{qn} - K_{rq})_+
\end{aligned}$$

Persamaan *spline truncated* multivariabel derajat 1 (linear) dengan  $n$  observasi dapat disajikan ke dalam bentuk matriks sebagai berikut:

$$\begin{aligned}
\begin{bmatrix} f_{p1} \\ f_{p2} \\ \vdots \\ f_{pn} \end{bmatrix} &= \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{q1} \\ 1 & x_{12} & x_{22} & \dots & x_{q2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{qn} \end{bmatrix} \begin{bmatrix} \delta_0^* \\ \delta_{11} \\ \vdots \\ \delta_{1q} \end{bmatrix} + \begin{bmatrix} (x_{11} - K_{11})_+ & \dots & (x_{11} - K_{r1})_+ \\ (x_{12} - K_{11})_+ & \dots & (x_{12} - K_{r1})_+ \\ \vdots & \ddots & \vdots \\ (x_{1n} - K_{11})_+ & \dots & (x_{1n} - K_{r1})_+ \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{31} \\ \vdots \\ \phi_{r1} \end{bmatrix} + \\
&\quad \begin{bmatrix} (x_{21} - K_{12})_+ & \dots & (x_{21} - K_{r2})_+ \\ (x_{22} - K_{12})_+ & \dots & (x_{22} - K_{r2})_+ \\ \vdots & \ddots & \vdots \\ (x_{2n} - K_{12})_+ & \dots & (x_{2n} - K_{r2})_+ \end{bmatrix} \begin{bmatrix} \phi_{22} \\ \phi_{32} \\ \vdots \\ \phi_{r2} \end{bmatrix} + \dots + \begin{bmatrix} (x_{q1} - K_{1q})_+ & \dots & (x_{q1} - K_{rq})_+ \\ (x_{q2} - K_{1q})_+ & \dots & (x_{q2} - K_{rq})_+ \\ \vdots & \ddots & \vdots \\ (x_{qn} - K_{1q})_+ & \dots & (x_{qn} - K_{rq})_+ \end{bmatrix} \begin{bmatrix} \phi_{2q} \\ \phi_{3q} \\ \vdots \\ \phi_{rq} \end{bmatrix}
\end{aligned}$$

Model matriks di atas dapat disajikan ke dalam bentuk Persamaan (4.4):

$$\sum_{p=1}^q \vec{f}_p(x_{pi}) = \mathbf{X}_0 \vec{\delta} + \mathbf{X}_1(\mathbf{K}_1) \vec{\phi}_1 + \mathbf{X}_2(\mathbf{K}_2) \vec{\phi}_2 + \dots + \mathbf{X}_q(\mathbf{K}_q) \vec{\phi}_q \quad (4.4)$$

dengan:

$$\vec{f}_p(x_{pi}) = \begin{bmatrix} f_{p1} \\ f_{p2} \\ \vdots \\ f_{pn} \end{bmatrix} \quad \mathbf{X}_0 = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{q1} \\ 1 & x_{12} & x_{22} & \dots & x_{q2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{qn} \end{bmatrix} \quad \vec{\delta} = \begin{bmatrix} \delta_0 \\ \delta_{11} \\ \vdots \\ \delta_{1q} \end{bmatrix}$$

$$\mathbf{X}_1(\mathbf{K}_1) = \begin{bmatrix} (x_{11} - K_{11})_+ & \dots & (x_{11} - K_{r1})_+ \\ (x_{12} - K_{11})_+ & \dots & (x_{12} - K_{r1})_+ \\ \vdots & \ddots & \vdots \\ (x_{1n} - K_{11})_+ & \dots & (x_{1n} - K_{r1})_+ \end{bmatrix} \quad \vec{\phi}_1 = \begin{bmatrix} \phi_{21} \\ \phi_{31} \\ \vdots \\ \phi_{r1} \end{bmatrix}$$

$$\mathbf{X}_2(\mathbf{K}_2) = \begin{bmatrix} (x_{21} - K_{12})_+ & \dots & (x_{21} - K_{r2})_+ \\ (x_{22} - K_{12})_+ & \dots & (x_{22} - K_{r2})_+ \\ \vdots & \ddots & \vdots \\ (x_{2n} - K_{12})_+ & \dots & (x_{2n} - K_{r2})_+ \end{bmatrix} \quad \vec{\phi}_2 = \begin{bmatrix} \phi_{22} \\ \phi_{32} \\ \vdots \\ \phi_{r2} \end{bmatrix}$$

hingga

$$\mathbf{X}_q(\mathbf{K}_q) = \begin{bmatrix} (x_{q1} - K_{1q})_+ & \dots & (x_{q1} - K_{rq})_+ \\ (x_{q2} - K_{1q})_+ & \dots & (x_{q2} - K_{rq})_+ \\ \vdots & \ddots & \vdots \\ (x_{qn} - K_{1q})_+ & \dots & (x_{qn} - K_{rq})_+ \end{bmatrix} \quad \vec{\phi}_q = \begin{bmatrix} \phi_{2q} \\ \phi_{3q} \\ \vdots \\ \phi_{rq} \end{bmatrix}$$

Vektor  $\vec{f}_p(x_{pi})$  berukuran  $(n \times 1)$ , vektor  $\vec{\delta}$  berukuran  $((q+1) \times 1)$ , vektor  $\vec{\phi}_1, \vec{\phi}_2$  hingga  $\vec{\phi}_q$  berukuran  $(r \times 1)$ . Matriks  $\mathbf{X}_0$  berukuran  $(n \times (q+1))$ , kemudian matriks  $\mathbf{X}_1(\mathbf{K}_1), \mathbf{X}_2(\mathbf{K}_2)$  hingga  $\mathbf{X}_q(\mathbf{K}_q)$  berukuran  $(n \times r)$ .

Selanjutnya berdasarkan Persamaan (4.4), maka:

$$\sum_{p=1}^q \vec{f}_p(x_{pi}) = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1(\mathbf{K}_1) & \dots & \mathbf{X}_q(\mathbf{K}_q) \end{bmatrix} \begin{bmatrix} \vec{\delta} \\ \vec{\phi}_1 \\ \vdots \\ \vec{\phi}_q \end{bmatrix}$$

Sehingga dapat dituliskan:

$$\sum_{p=1}^q \vec{f}_p(x_{pi}) = \mathbf{X}(\phi) \vec{\beta} \quad (4.5)$$

dengan

$$\mathbf{X}(\phi) = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1(\mathbf{K}_1) & \cdots & \mathbf{X}_q(\mathbf{K}_q) \end{bmatrix} \text{ dan } \vec{\beta} = \begin{bmatrix} \vec{\delta} \\ \vec{\phi}_1 \\ \vdots \\ \vec{\phi}_q \end{bmatrix}$$

Selanjutnya proses estimasi parameter dapat menggunakan OLS dengan cara meminimumkan jumlah kuadrat *error*.

#### 4.1.2 Estimator Kurva Regresi untuk Fungsi Kernel Gaussian

Misalkan diberikan data berpasangan  $t_i$  dan  $y_i$ , dimana  $i=1,2,\dots,n$  yang mengikuti model regresi nonparametrik. Dengan  $y_i$  adalah variabel respon, sedangkan  $t_i$  adalah variabel prediktor. Hubungan  $t_i$  dan  $y_i$  dapat dimodelkan secara fungsional ke dalam bentuk:

$$y_i = h(t_i) + \varepsilon_i; i = 1, 2, \dots, n \quad (4.6)$$

Kurva regresi dari  $h(t_i)$  merupakan kurva yang tidak diketahui bentuknya. Selanjutnya kurva regresi  $h(t_i)$  akan dihipotesiskan dengan estimasi kurva regresi yang disajikan seperti pada Persamaan (4.7).

$$\begin{aligned} \hat{h}_\alpha(t) &= n^{-1} \sum_{i=1}^n \left[ \frac{K_\alpha(t-t_i)}{n^{-1} \sum_{i=1}^n K_\alpha(t-t_i)} \right] y_i \\ &= n^{-1} \left[ \sum_{i=1}^n W_{ai}(t) \right] y_i \end{aligned} \quad (4.7)$$

dengan:

$$W_{ai}(t) = \frac{K_\alpha(t-t_i)}{n^{-1} \sum_{i=1}^n K_\alpha(t-t_i)} \text{ dan } K_\alpha(t-t_i) = \frac{1}{\alpha} K\left(\frac{t-t_i}{\alpha}\right)$$

Pada penelitian ini, fungsi kernel yang digunakan adalah kernel Gaussian seperti pada Persamaan (4.8).

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}(-z^2)\right); I_{[-\infty, \infty]}(z) \quad (4.8)$$

Estimator kernel sangat bergantung pada fungsi kernel  $K$  seperti pada Persamaan (4.8) dan *bandwidth*. Berdasarkan persamaan Kernel yang berlaku untuk setiap  $t = t_1, t = t_2, \dots, t = t_n$ , maka dapat dituliskan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} \hat{h}_\alpha(t_1) \\ \hat{h}_\alpha(t_2) \\ \vdots \\ \hat{h}_\alpha(t_n) \end{bmatrix} = \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\alpha i}(t_1) y_i \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_n) y_i \end{bmatrix}$$

Dapat dijabarkan sebagai contoh:

$$n^{-1} \sum_{i=1}^n W_{\alpha i}(t_1) y_i = n^{-1} W_{\alpha 1}(t_1) y_1 + n^{-1} W_{\alpha 2}(t_1) y_2 + \dots + n^{-1} W_{\alpha n}(t_1) y_n$$

$$n^{-1} \sum_{i=1}^n W_{\alpha i}(t_2) y_i = n^{-1} W_{\alpha 1}(t_2) y_1 + n^{-1} W_{\alpha 2}(t_2) y_2 + \dots + n^{-1} W_{\alpha n}(t_2) y_n$$

⋮

$$n^{-1} \sum_{i=1}^n W_{\alpha i}(t_n) y_i = n^{-1} W_{\alpha 1}(t_n) y_1 + n^{-1} W_{\alpha 2}(t_n) y_2 + \dots + n^{-1} W_{\alpha n}(t_n) y_n$$

Maka,

$$\begin{aligned} \begin{bmatrix} \hat{h}_\alpha(t_1) \\ \hat{h}_\alpha(t_2) \\ \vdots \\ \hat{h}_\alpha(t_n) \end{bmatrix} &= \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\alpha i}(t_1) y_i \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(t_n) y_i \end{bmatrix} = \begin{bmatrix} n^{-1} W_{\alpha 1}(t_1) y_1 + n^{-1} W_{\alpha 2}(t_1) y_2 + \dots + n^{-1} W_{\alpha n}(t_1) y_n \\ n^{-1} W_{\alpha 1}(t_2) y_1 + n^{-1} W_{\alpha 2}(t_2) y_2 + \dots + n^{-1} W_{\alpha n}(t_2) y_n \\ \vdots \\ n^{-1} W_{\alpha 1}(t_n) y_1 + n^{-1} W_{\alpha 2}(t_n) y_2 + \dots + n^{-1} W_{\alpha n}(t_n) y_n \end{bmatrix} \\ &= \begin{bmatrix} n^{-1} W_{\alpha 1}(t_1) & n^{-1} W_{\alpha 2}(t_1) & \cdots & n^{-1} W_{\alpha n}(t_1) \\ n^{-1} W_{\alpha 1}(t_2) & n^{-1} W_{\alpha 2}(t_2) & \cdots & n^{-1} W_{\alpha n}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\alpha 1}(t_n) & n^{-1} W_{\alpha 2}(t_n) & \cdots & n^{-1} W_{\alpha n}(t_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \mathbf{D}(\alpha) \vec{Y} \end{aligned} \quad (4.9)$$



Berdasarkan Persamaan (4.9) maka estimator kernel dapat dituliskan  $\hat{h}_\alpha(t) = \mathbf{D}(\alpha)\vec{Y}$ . Vektor  $\hat{h}_\alpha(t)$  berukuran  $(n \times 1)$ , vektor  $\vec{Y}$  berukuran  $(n \times 1)$  dan matriks  $\mathbf{D}(\alpha)$  berukuran  $(n \times n)$ .

#### 4.1.3 Estimator Campuran Fungsi *Spline Truncated* Multivariabel dan Kernel Gaussian

Langkah awal adalah mendefinisikan model regresi nonparametrik *spline truncated* dan kernel Gaussian. Diberikan data berpasangan dengan variabel prediktor sebanyak  $q$  untuk komponen *spline truncated* dan hanya 1 untuk komponen kernel sedemikian sehingga dapat ditulis  $(x_{1i}, x_{2i}, \dots, x_{qi}, t_i, y_i)$  dengan  $i = 1, 2, \dots, n$ .

Asumsikan bahwa pola hubungan antara variabel prediktor  $(x_{1i}, x_{2i}, \dots, x_{qi}, t_i)$  dan variabel respon  $y_i$  mengikuti model regresi nonparametrik:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{qi}) + h(t_i) + \varepsilon_i \quad (4.10)$$

Kurva regresi dari komponen *spline truncated* yaitu  $f(x_{1i}, x_{2i}, \dots, x_{qi})$  diasumsikan bersifat aditif, sehingga didapatkan:

$$f(x_{1i}, x_{2i}, \dots, x_{qi}) = \sum_{p=1}^q f_p(x_{pi}), \quad i = 1, 2, \dots, n$$

Selanjutnya model regresi nonparametrik *spline truncated* dan kernel Gaussian dapat dituliskan:

$$y_i = \sum_{p=1}^q f_p(x_{pi}) + h(t_i) + \varepsilon_i \quad \text{dengan } i = 1, 2, \dots, n \quad (4.11)$$

Berdasarkan Persamaan (4.11), dan bentuk dari masing-masing komponen yang telah di dapatkan sebelumnya. Maka dapat dituliskan:

$$y_i = \mu(x_i, t_i) + \varepsilon_i \quad (4.12)$$

dimana  $\mu(x_i, t_i) = \sum_{p=1}^q f_p(x_{pi}) + h_\alpha(t_i)$ . Sehingga apabila dilakukan substitusi akan

didapatkan  $y_i = \left[ \sum_{p=1}^q f_p(x_{pi}) + h_\alpha(t_i) \right] + \varepsilon_i$ .

Kemudian dapat di sajikan dalam bentuk matriks menjadi:

$$\vec{Y} = \mathbf{X}(\phi)\vec{\beta} + \mathbf{D}(\alpha)\vec{Y} + \vec{\varepsilon} \quad (4.13)$$

Estimasi parameter dari  $\vec{\beta}$  dapat diperoleh dengan menggunakan metode *Ordinary Least Squares* (OLS) berdasarkan Persamaan (4.13) di atas. Dengan menggunakan metode OLS yaitu suatu metode untuk mendapatkan estimator dengan cara meminimumkan jumlah kuadrat *error* sedemikian sehingga *error* nya dapat disajikan:

$$\begin{aligned} \vec{\varepsilon} &= \vec{Y} - [\mathbf{X}(\phi)\vec{\beta} + \mathbf{D}(\alpha)\vec{Y}] \\ &= [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} - \mathbf{X}(\phi)\vec{\beta} \end{aligned} \quad (4.14)$$

Selanjutnya jumlah kuadrat *error* bisa didapatkan sebagai berikut:

$$\begin{aligned} Q(\vec{\beta} | \phi, \alpha) &= \sum_{i=1}^n \vec{\varepsilon}_i^2 = \vec{\varepsilon}^T \vec{\varepsilon} \\ &= ([\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} - \mathbf{X}(\phi)\vec{\beta})^T ([\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} - \mathbf{X}(\phi)\vec{\beta}) \\ &= (\vec{Y}^T [\mathbf{I} - \mathbf{D}(\alpha)]^T - \vec{\beta}^T \mathbf{X}(\phi)^T) ([\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} - \mathbf{X}(\phi)\vec{\beta}) \\ &= \left\| [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} \right\|^2 - \vec{Y}^T [\mathbf{I} - \mathbf{D}(\alpha)]^T \mathbf{X}(\phi)\vec{\beta} - \vec{\beta}^T \mathbf{X}(\phi)^T \\ &\quad [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi)\vec{\beta} \\ &= \left\| [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} \right\|^2 - \vec{\beta}^T \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} - \vec{\beta}^T \mathbf{X}(\phi)^T \\ &\quad [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi)\vec{\beta} \\ &= \left\| [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} \right\|^2 - 2\vec{\beta}^T \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi)\vec{\beta} \end{aligned} \quad (4.15)$$

Untuk mendapatkan estimasi dari  $\vec{\beta}$ , maka dapat melakukan derivatif parsial dari persamaan diatas terhadap  $\vec{\beta}$  sebagai berikut:

$$\begin{aligned} \frac{\partial Q(\vec{\beta} | \phi, \alpha)}{\partial \vec{\beta}} &= \frac{\partial \left( \left\| [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} \right\|^2 - 2\vec{\beta}^T \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} + \vec{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi)\vec{\beta} \right)}{\partial \vec{\beta}} \\ &= -2\mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y} + 2\mathbf{X}(\phi)^T \mathbf{X}(\phi)\vec{\beta} \end{aligned}$$

Apabila persamaan diatas, disamakan dengan nol dengan:

$$\frac{\partial Q(\vec{\beta} | \phi, \alpha)}{\partial \vec{\beta}} = \vec{0} \quad (4.16)$$

Maka akan diperoleh:

$$\mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta} - \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} = \vec{0}$$

Selanjutnya, dilakukan pemindahan ruas dari persamaan atas, sehingga dapat ditulis

$$\mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta} = \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} \quad (4.17)$$

Maka akan di dapatkan estimator untuk  $\hat{\beta}$  sebagai berikut:

$$\hat{\beta} = [\mathbf{X}(\phi)^T \mathbf{X}(\phi)]^{-1} \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} \quad (4.18)$$

Dapat di ringkas menjadi:

$$\hat{\beta} = \mathbf{C}(\phi, \alpha) \vec{Y} \quad (4.19)$$

dengan  $\mathbf{C} = [\mathbf{X}(\phi)^T \mathbf{X}(\phi)]^{-1} \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]$ .

Sehingga didapatkan estimator kurva *spline truncated* yang diketahui sebelumnya

adalah  $\sum_{p=1}^q \hat{f}_p(x_{pi}) = \mathbf{X}(\phi) \hat{\beta}$ , sehingga dapat di tuliskan:

$$\begin{aligned} \sum_{p=1}^q \hat{f}_p(x_p) &= \mathbf{X}(\phi) \hat{\beta} \\ &= \mathbf{X}(\phi) [\mathbf{X}(\phi)^T \mathbf{X}(\phi)]^{-1} \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)] \vec{Y} \end{aligned} \quad (4.20)$$

maka  $\sum_{p=1}^q \hat{f}_p(x_{pi}) = \mathbf{A}(\phi, \alpha) \vec{Y}$ .

dengan matriks  $\mathbf{A}(\phi, \alpha)$  adalah  $\mathbf{A}(\phi, \alpha) = \mathbf{X}(\phi) [\mathbf{X}(\phi)^T \mathbf{X}(\phi)]^{-1} \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]$

Selanjutnya bentuk estimator persamaan kernel Gaussian sebelumnya diketahui

$$\hat{h}_\alpha(t) = \mathbf{D}(\alpha) \vec{Y} \quad (4.21)$$

Maka akan di dapatkan estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian sebagai berikut:

$$\begin{aligned} \hat{\mu}_{\phi, \alpha}(x, t) &= \sum_{p=1}^q \hat{f}_p(x_{pi}) + \hat{h}_\alpha(t) \\ &= \mathbf{A}(\phi, \alpha) \vec{Y} + \mathbf{D}(\alpha) \vec{Y} \\ &= \mathbf{B}(\phi, \alpha) \vec{Y} \end{aligned} \quad (4.22)$$

Dengan matriks  $\mathbf{B}(\phi, \alpha)$  adalah  $[\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]$ .

Matriks  $\mathbf{B}(\phi, \alpha)\bar{Y}$  sangat tergantung pada  $\mathbf{A}(\phi, \alpha)$  yang merupakan komponen *spline truncated* dengan titik knot  $\phi = (K_1, K_2, \dots, K_r)^T$  dan  $\mathbf{D}(\alpha)$  yang merupakan komponen fungsi kernel Gaussian dengan parameter  $\alpha$  yang merupakan *bandwidth*.

Untuk memperoleh estimator campuran regresi nonparametrik *spline truncated* dan kernel Gaussian perlu dilakukan pemilihan titik knot dan parameter *bandwidth* yang optimal. Metode yang digunakan pada penelitian ini adalah *Cross-Validation* (CV), *Generalized Cross-Validation* (GCV) dan *Unbiased Risk* (UBR).

#### 4.1.4 Cross-Validation (CV)

*Cross Validation* (CV) merupakan suatu metode yang dikembangkan oleh Craven dan Wahba (1979). Formula metode CV yang dikembangkan oleh Craven dan Wahba masih terbatas pada bentuk estimator tunggal, dimana dapat dituliskan sebagai berikut:

$$CV(k) = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(y_i - \hat{y}_i)}{(1 - \mathbf{A}_{ii}(k))} \right]^2 \quad (4.23)$$

Dimana  $\mathbf{A}(k)$  adalah  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ .

Kelebihan dari Metode CV adalah tidak memerlukan informasi varians  $\sigma^2$ . Selanjutnya Metode CV dapat juga digunakan pada bentuk estimator campuran. Formula metode CV yang telah dimodifikasi untuk bentuk estimator campuran dapat dituliskan seperti pada Persamaan (4.24).

$$\begin{aligned} CV(\phi_{opt}, \alpha_{opt}) &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{(y_i - \hat{y}_i)}{(1 - [\mathbf{A}_{ii}(\phi, \alpha) + \mathbf{D}_{ii}(\alpha)])} \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{(y_i - \hat{y}_i)}{(1 - [\mathbf{B}_{ii}(\phi, \alpha)])} \right]^2 \end{aligned} \quad (4.24)$$

dimana:

- $y_i$  : variabel respon ke- $i$
- $\hat{y}_i$  : taksiran variabel respon ke- $i$

Formula metode CV pada Persamaan (4.24) merupakan modifikasi yang disesuaikan untuk model estimator campuran. Selanjutnya dapat diketahui jika metode CV memberikan bobot yang berbeda-beda pada setiap observasinya sesuai dengan kontribusinya. Matriks  $\mathbf{B}(\phi, \alpha)$  dapat dicari berdasarkan Persamaan (4.22), dimana matriks  $\mathbf{B}(\phi, \alpha)$  sangat tergantung pada matriks komponen *spline truncated* dan kernel Gaussian. Taksiran variabel respon ke- $i$  ( $\hat{y}_i$ ) didapatkan dari hasil pemodelan regresi nonparametrik dengan estimator campuran. Nilai CV terkecil yang dihasilkan akan memberikan titik knot dan *bandwidth* yang optimal.

#### 4.1.5 Generalized Cross-Validation (GCV)

Metode GCV merupakan generalisasi dari metode CV. Metode GCV dikembangkan oleh Wahba (1990). Formula metode GCV yang dikembangkan oleh Wahba (1990) masih terbatas pada bentuk estimator tunggal, dimana dapat dituliskan sebagai berikut:

$$GCV(k) = \frac{MSE(k)}{\left(n^{-1} \text{trace}\{\mathbf{I} - \mathbf{A}(k)\}\right)^2} \quad (4.25)$$

Dimana  $\mathbf{A}(k)$  adalah  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$

dan  $MSE(k)$  yaitu  $n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ .

Hal yang sama seperti metode CV, bahwa metode GCV tidak memerlukan informasi varians  $\sigma^2$ . Metode GCV dapat digunakan pada bentuk estimator campuran. Formula metode GCV yang telah dimodifikasi untuk bentuk estimator campuran dapat dituliskan seperti pada Persamaan (4.26).

$$GCV(\phi_{opt}, \alpha_{opt}) = \left[ \frac{MSE(\phi, \alpha)}{\left(n^{-1} \text{trace}\{\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]\}\right)^2} \right] \quad (4.26)$$

dimana:

$$MSE(\phi, \alpha) \quad : \quad n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$\mathbf{I}$  : matriks identitas

Dapat dijabarkan dalam bentuk lain Persamaan (4.26) seperti pada Persamaan (4.27):

$$\begin{aligned}
GCV(\phi_{opt}, \alpha_{opt}) &= \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left[ n^{-1} \text{trace} \{ \mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)] \} \right]^2} \\
&= \frac{n^{-1} \left[ \bar{\mathbf{Y}}^T (\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)])^T (\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]) \bar{\mathbf{Y}} \right]}{\left[ n^{-1} \text{trace} \{ \mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)] \} \right]^2} \quad (4.27) \\
&= \frac{n^{-1} \left[ \bar{\mathbf{Y}}^T (\mathbf{I} - [\mathbf{B}(\phi, \alpha)])^T (\mathbf{I} - [\mathbf{B}(\phi, \alpha)]) \bar{\mathbf{Y}} \right]}{\left[ n^{-1} \text{trace} \{ \mathbf{I} - [\mathbf{B}(\phi, \alpha)] \} \right]^2}
\end{aligned}$$

Formula metode GCV pada Persamaan (4.27) merupakan modifikasi yang disesuaikan untuk model estimator campuran. Metode GCV merupakan metode CV yang terboboti, yaitu dengan memberikan bobot yang sama pada setiap observasinya. Nilai GCV terkecil yang dihasilkan akan memberikan titik knot dan *bandwidth* yang optimal.

#### 4.1.6 Unbiased Risk (UBR)

Metode UBR dapat digunakan untuk menentukan parameter penghalus yang optimal ketika ada informasi mengenai  $\sigma^2$  atau  $\sigma^2$  telah diketahui. Metode ini diperkenalkan oleh Wang (1997) yang terbatas pada bentuk estimator tunggal. Formula dari metode UBR dapat dituliskan sebagai berikut:

$$U(k) = \frac{1}{n} \left\{ \begin{aligned} &\|(\mathbf{I} - \mathbf{A}(k)) \bar{\mathbf{Y}}\|^2 + \frac{\sigma^2}{n} \text{trace}[\mathbf{I} - \mathbf{A}(k)]^2 + \\ &\frac{\sigma^2}{n} \text{trace}[\mathbf{A}^2(k)]^2 \end{aligned} \right\} \quad (4.28)$$

dimana taksiran dari  $\sigma^2$  seperti pada Persamaan (4.29).

$$\sigma^2 = \frac{\|(\mathbf{I} - \mathbf{A}(k)) \bar{\mathbf{Y}}\|^2}{\text{trace}[(\mathbf{I} - \mathbf{A}(k)) \bar{\mathbf{Y}}]} \quad (4.29)$$

Selanjutnya Persamaan (4.28) dan (4.29) dapat dimodifikasi untuk bentuk estimator campuran. Formula metode UBR yang telah dimodifikasi untuk bentuk estimator campuran dapat dituliskan pada Persamaan (4.30).

$$\begin{aligned}
UBR(\phi_{opt}, \alpha_{opt}) &= n^{-1} \left\{ \begin{aligned} &\|(\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)])\bar{Y}\|^2 + \\ &\frac{\hat{\sigma}^2}{n} \text{trace}[\mathbf{I} - [\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]]^2 + \\ &\frac{\hat{\sigma}^2}{n} \text{trace}[\mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha)]^2 \end{aligned} \right\} \\
&= n^{-1} \left\{ \begin{aligned} &\|(\mathbf{I} - \mathbf{B}(\phi, \alpha))\bar{Y}\|^2 + \frac{\hat{\sigma}^2}{n} \text{trace}[\mathbf{I} - \mathbf{B}(\phi, \alpha)]^2 + \\ &\frac{\hat{\sigma}^2}{n} \text{trace}[\mathbf{B}(\phi, \alpha)^2] \end{aligned} \right\}
\end{aligned} \tag{4.30}$$

Estimasi dari  $\sigma^2$  dapat diperoleh dengan formula sebagai berikut:

$$\hat{\sigma}^2 = \frac{\|(\mathbf{I} - \mathbf{B}(\phi, \alpha))\bar{Y}\|^2}{\text{trace}[(\mathbf{I} - \mathbf{B}(\phi, \alpha))\bar{Y}]} \tag{4.31}$$

Nilai UBR terkecil yang dihasilkan akan memberikan titik knot dan *bandwidth* yang optimal. Matriks  $\mathbf{B}(\phi, \alpha)$  dapat dicari berdasarkan Persamaan (4.22), dimana matriks  $\mathbf{B}(\phi, \alpha)$  sangat tergantung pada matriks komponen *spline truncated* dan kernel Gaussian.

## 4.2 Kajian Simulasi

Berdasarkan tujuan penelitian kedua, maka akan dilakukan kajian bagaimana hasil perbandingan kinerja metode pemilihan titik knot dan *bandwidth* menggunakan CV, GCV dan UBR pada data simulasi.

Fungsi yang dirancang untuk studi simulasi ini mengikuti model regresi nonparametrik estimator campuran:

$$y_i = \mu(x_i, t_i) + \varepsilon_i, \text{ dengan } \mu(x_i, t_i) = f(x_i) + h(t_i) \tag{4.32}$$

$f(x_i)$  adalah fungsi *smooth* yang di definisikan sebagai komponen *spline truncated*

dengan  $f(x_i) = \frac{\sin(\pi x_i^2)^4}{\sin(x_i^2)}$ . Kemudian  $h(t_i)$  adalah fungsi yang di definisikan

sebagai komponen kernel dengan  $h(t_i) = t_i^2 \sin t_i$ .

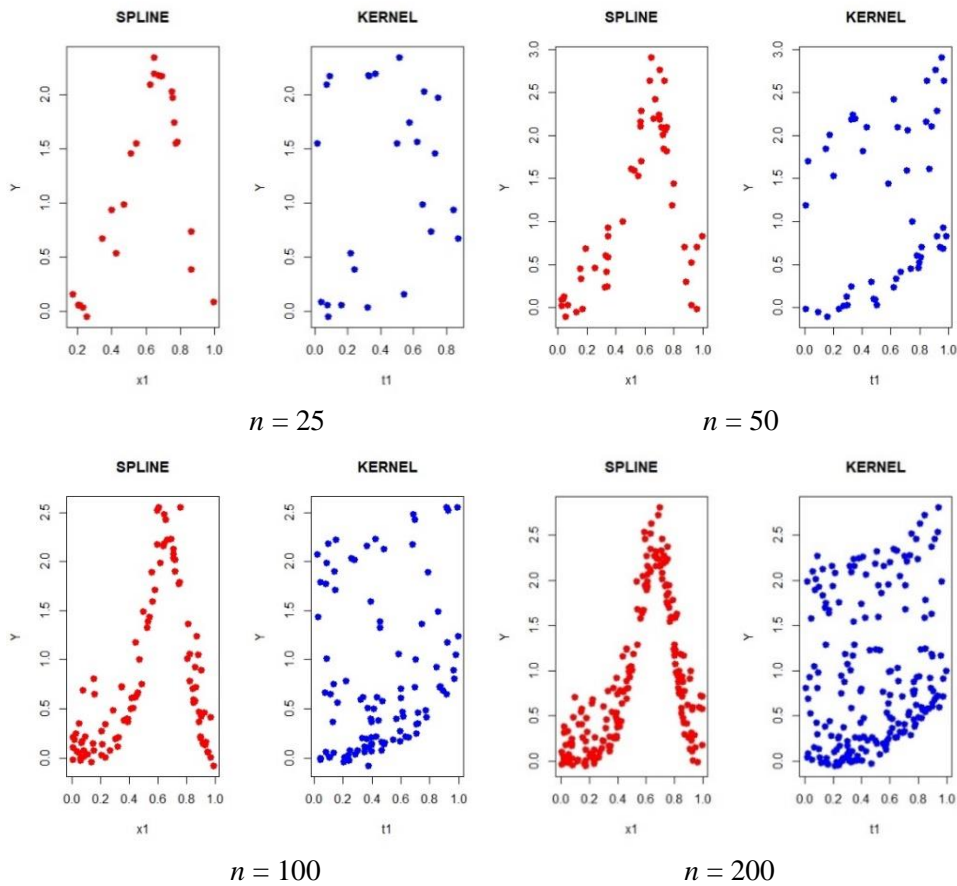
Variabel  $x_i$  yang di tetapkan untuk komponen *spline truncated* dan  $t_i$  yang di tetapkan untuk komponen kernel dibangkitkan secara independen dari distribusi

Uniform  $U(0,1)$ . Kemudian dibangkitkan *error* atau sisaan yaitu  $\varepsilon_i$ , dimana  $i = 1, 2, \dots, n$  yang mengikuti distribusi Normal  $(0, \sigma^2)$ , sehingga  $\varepsilon_i \sim N(0, \sigma^2)$ . Maka akan didapatkan variabel respon  $y_i$  dari model regresi nonparametrik estimator campuran sebagai berikut:

$$y_i = f(x_i) + h(t_i) + \varepsilon_i \rightarrow \left[ \frac{\sin(\pi x_i^2)^4}{\sin(x_i^2)} \right] + [t_i^2 \sin t_i] + \varepsilon_i \quad (4.33)$$

Simulasi dilakukan pada berbagai kondisi yang di atur. Pada penelitian ini variasi ukuran sampel dengan  $n$  yang akan dicobakan adalah 25, 50, 100 dan 200. Kombinasi dari varians *error*  $\sigma^2$  yang akan dicobakan adalah 0,05; 0,5 dan 1.

Sebagai ilustrasi, ditampilkan diagram pencar antara variabel respon  $y_i$  dengan masing-masing variabel prediktor  $(x_i, t_i)$  dengan  $\sigma^2 = 0,05$  dan jumlah sampel yang bervariasi ditampilkan pada Gambar 4.1.



**Gambar 4.1** Diagram pencar antara variabel prediktor dengan respon data simulasi



Gambar 4.1 menunjukkan bahwa masing-masing dari variabel prediktor memiliki bentuk kurva regresi yang berbeda. Pada variabel  $x_1$  menunjukkan karakteristik dari estimator *spline truncated*, dimana memiliki pola data yang berubah-ubah pada sub-sub interval tertentu. Sedangkan hal yang berbeda terjadi pada variabel  $t_1$ , dimana menunjukkan pola data yang tidak mempunyai pola tertentu, sehingga akan di dekati dengan bentuk estimator kernel.

Pada kajian simulasi ini, banyaknya titik knot yang akan dicobakan hanya 1 titik knot, berlaku untuk variabel yang di definisikan sebagai komponen *spline truncated*. Replikasi untuk setiap data bangkitan sebanyak 20 kali.

Hasil simulasi berupa rata-rata nilai *Cross-Validation* (CV) dan koefisien determinasi ( $R^2$ ) dari setiap replikasi atau perulangan yang didapatkan berdasarkan hasil pemodelan regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian di sajikan pada Tabel 4.1.

**Tabel 4.1** Hasil Simulasi dengan Metode *Cross-Validation* (CV)

Varians	Rata-Rata	Ukuran Sampel			
		$n=25$	$n=50$	$n=100$	$n=200$
$\sigma^2 = 0,05$	CV	0,136	0,120	0,110	0,109
	$R^2$	85,65%	84,75%	84,33%	84,45%
$\sigma^2 = 0,5$	CV	0,393	0,374	0,371	0,364
	$R^2$	66,38%	65,12%	61,84%	61,02%
$\sigma^2 = 1$	CV	1,128	1,133	1,133	1,087
	$R^2$	40,86%	37,42%	38,05%	35,41%

Berdasarkan Tabel 4.1, dapat diketahui jika semakin meningkatnya nilai varians yang dicobakan, maka terjadi kecenderungan untuk menurunkan nilai  $R^2$  untuk setiap variasi ukuran sampel yang dicobakan. Adapun setelah dilakukan pemeriksaan menggunakan statistik uji Kolmogorov-Smirnov, didapatkan hasil jika residual dari setiap hasil pemodelan untuk setiap kombinasi variasi ukuran sampel dan varians mengikuti distribusi Normal.

Selanjutnya hasil simulasi berupa rata-rata nilai *Generalized Cross-Validation* (GCV) dan koefisien determinasi ( $R^2$ ) dari setiap replikasi atau

perulangan yang didapatkan berdasarkan hasil pemodelan regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian di sajikan pada Tabel 4.2.

**Tabel 4.2** Hasil Simulasi dengan Metode *Generalized Cross-Validation* (GCV)

Varians	Rata-Rata	Ukuran Sampel			
		<i>n=25</i>	<i>n=50</i>	<i>n=100</i>	<i>n=200</i>
$\sigma^2 = 0,05$	GCV	1,473	2,733	5,010	8,403
	R <sup>2</sup>	86,48%	85,30%	85,26%	85,34%
$\sigma^2 = 0,5$	GCV	1,475	3,197	4,436	10,153
	R <sup>2</sup>	71,26%	67,25%	63,38%	61,89%
$\sigma^2 = 1$	GCV	2,168	4,033	8,096	15,192
	R <sup>2</sup>	56,70%	44,86%	41,18%	37,15%

Berdasarkan Tabel 4.2, dapat diketahui jika hal yang sama terjadi, dimana semakin meningkatnya nilai varians yang dicobakan, maka terjadi kecenderungan untuk menurunkan nilai R<sup>2</sup> untuk setiap variasi ukuran sampel yang dicobakan. Pada kasus ini, setelah dilakukan pemeriksaan didapatkan jika residual dari setiap hasil pemodelan untuk setiap kombinasi variasi ukuran sampel dan varians mengikuti distribusi Normal.

Hasil simulasi berupa rata-rata nilai *Unbiased Risk* (UBR) dan koefisien determinasi (R<sup>2</sup>) dari setiap replikasi atau perulangan yang didapatkan berdasarkan hasil pemodelan regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian di sajikan pada Tabel 4.3.

**Tabel 4.3** Hasil Simulasi dengan Metode *Unbiased Risk* (UBR)

Varians	Rata-Rata	Ukuran Sampel			
		<i>n=25</i>	<i>n=50</i>	<i>n=100</i>	<i>n=200</i>
$\sigma^2 = 0,05$	UBR	0,009	0,008	0,004	0,004
	R <sup>2</sup>	83,71%	82,74%	82,72%	82,59%
$\sigma^2 = 0,5$	UBR	0,014	0,007	0,005	0,004
	R <sup>2</sup>	64,23%	62,38%	59,42%	59,78%
$\sigma^2 = 1$	UBR	0,016	0,009	0,006	0,004
	R <sup>2</sup>	38,28%	35,75%	36,19%	34,04%

Berdasarkan Tabel 4.3, dapat diketahui dimana semakin meningkatnya nilai varians yang dicobakan, maka terjadi kecenderungan untuk menurunkan nilai R<sup>2</sup> untuk

setiap variasi ukuran sampel yang dicobakan. Setelah dilakukan pemeriksaan, didapatkan hasil jika residual dari setiap hasil pemodelan untuk setiap kombinasi variasi ukuran sampel dan varians mengikuti distribusi Normal.

Tabel 4.1, 4.2 dan 4.3 menunjukkan hasil dari rata-rata nilai CV, GCV dan UBR serta nilai koefisien determinasi ( $R^2$ ) untuk setiap metode penentuan parameter penghalus yang optimal (titik knot dan *bandwidth*) pada model regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian.

Untuk ukuran sampel  $n$  yang dibuat bervariasi yaitu 25, 50, 100, 200 dengan semua variasi dari varians yang dicobakan, metode GCV memberikan hasil estimasi titik knot dan *bandwidth* yang baik dibandingkan metode CV dan UBR. Hal ini ditunjukkan dari nilai koefisien determinasi ( $R^2$ ) yang didapatkan dari setiap percobaan dengan metode GCV, lebih tinggi dibandingkan kedua metode lainnya.

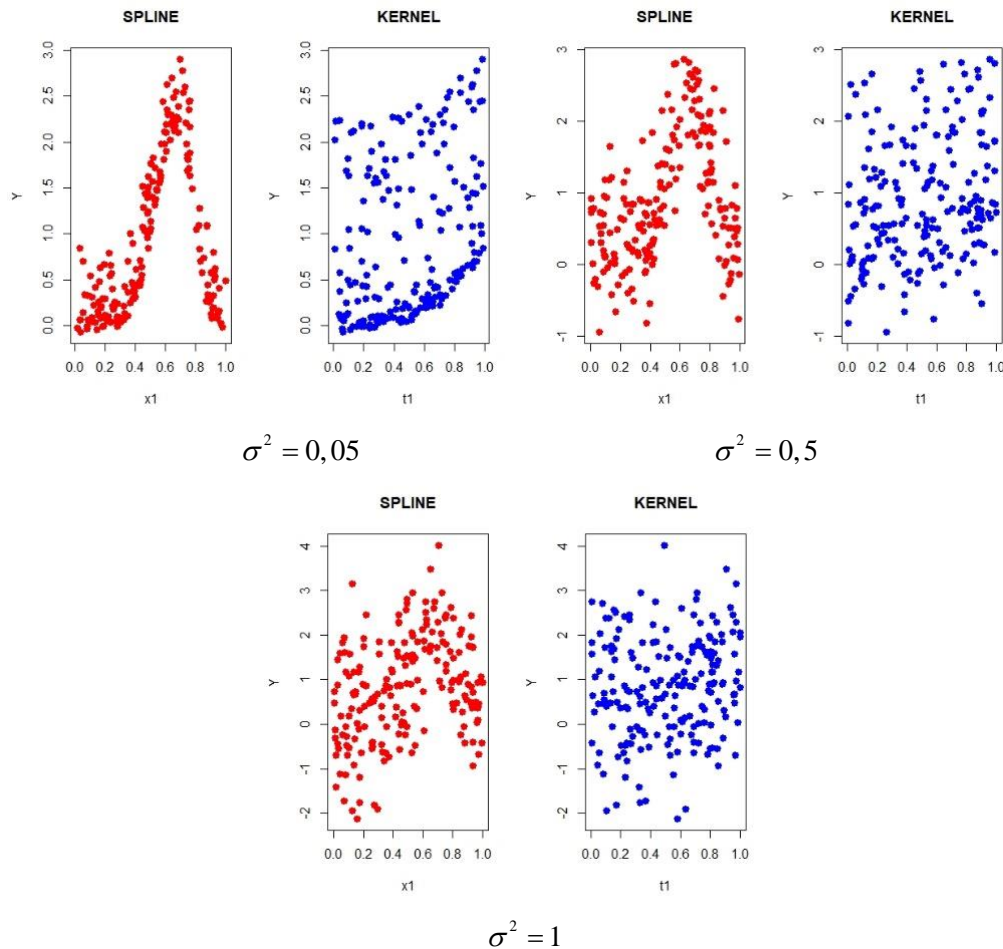
Sebagai contoh untuk jumlah sampel  $n=25$  dan varians dari *error* nya adalah  $\sigma^2 = 0,05$ , dengan menggunakan metode GCV dalam proses pemilihan titik knot dan *bandwidth* optimal diperoleh rata-rata nilai GCV adalah sebesar 1,473 dengan nilai  $R^2$  sebesar 86,48%. Sedangkan dengan menggunakan metode CV serta kondisi yang sama, diperoleh rata-rata nilai CV sebesar 0,136 dan nilai  $R^2$  yaitu 85,65%. Penggunaan metode UBR dengan kondisi yang sama, didapatkan rata-rata nilai UBR sebesar 0,009 dan nilai  $R^2$  sebesar 83,71%. Selanjutnya yang membedakan antara metode CV, GCV dengan UBR adalah pada metode UBR perlu adanya taksiran varians.

Dampak dari ukuran varians  $\sigma^2$  yang dibuat bervariasi pada penelitian ini berpengaruh terhadap hasil simulasi. Varians yang dicobakan adalah 0,05; 0,5 dan 1. Terlihat bahwa semakin meningkat nilai varians yang dicobakan maka terjadi kecenderungan penurunan besarnya nilai  $R^2$  untuk semua metode yang digunakan baik CV, GCV dan UBR.

Sebagai contoh lainnya jika jumlah sampel  $n=200$  dan varians dari *error* nya adalah  $\sigma^2 = 1$ , dengan menggunakan metode GCV dalam proses pemilihan titik knot dan *bandwidth* optimal diperoleh rata-rata nilai GCV adalah sebesar 15,192 dengan nilai  $R^2$  sebesar 37,15%. Sedangkan dengan menggunakan metode CV serta kondisi yang sama, diperoleh rata-rata nilai CV sebesar 1,087 dan nilai  $R^2$  yaitu

35,41%. Penggunaan metode UBR dengan kondisi yang sama, didapatkan rata-rata nilai UBR sebesar 0,004 dan nilai  $R^2$  sebesar 34,04%.

Hal yang terjadi selanjutnya adalah semakin bertambahnya ukuran sampel  $n$  yang dicobakan dan juga meningkatnya ukuran varians, maka nilai  $R^2$  yang didapatkan untuk semua metode cenderung menurun. Dampak dari varians ini sangat krusial dan penting dalam prosesnya. Varians menunjukkan penyimpangan data dari rata-ratanya, sehingga semakin tinggi nilai varians yang dicobakan, maka akan ada kecenderungan data yang menyebar jauh dari nilai rata-ratanya. Ilustrasi bangkitan data dengan  $n = 200$  dan berbagai kondisi varians ditampilkan pada Gambar 4.2.



**Gambar 4.2** Ilustrasi dampak dari varians pada pola data

Berdasarkan Gambar 4.2, dapat dilihat pada varians  $\sigma^2 = 0,05$ , untuk variabel yang di definisikan sebagai komponen *spline truncated* terlihat jelas masih menunjukkan

pola yang berubah-ubah pada sub-sub interval tertentu, sedangkan untuk komponen kernel terlihat tidak memiliki pola tertentu. Semakin meningkatnya nilai varians yang dicobakan, sebagai contoh pada  $\sigma^2 = 1$ , maka pola data terlihat sangat menyebar. Pada varians  $\sigma^2 = 1$ , untuk komponen *spline truncated* terlihat secara implisit menunjukkan pola yang berubah-ubah pada sub-sub interval tertentu, akan tetapi ada kecenderungan pola yang menyebar, sedangkan untuk komponen kernel terlihat semakin menyebar dan tidak memiliki pola.

Berdasarkan dampak dari ukuran varians dan ukuran sampel yang dicobakan, dapat diketahui jika metode GCV tetap memberikan hasil estimasi titik knot dan *bandwidth* yang optimal sehingga menghasilkan nilai koefisien determinasi ( $R^2$ ) yang tertinggi dibandingkan kedua metode lainnya yang dicobakan untuk setiap kondisi.

Dari hasil studi simulasi, dapat diketahui secara keseluruhan bahwa hasil estimasi titik knot dan *bandwidth* yang dihasilkan dari metode CV, GCV dan UBR dapat dikatakan cukup baik. Namun metode GCV memberikan performa hasil serta ketepatan yang lebih baik untuk setiap kombinasi jumlah sampel dan variasi varians yang dicobakan. Metode GCV menghasilkan titik knot dan *bandwidth* yang optimal sehingga didapatkan nilai koefisien determinasi yang terbesar untuk setiap kombinasi.

Sehingga berdasarkan hasil analisis kajian simulasi ini, diketahui bahwa metode GCV lebih sesuai untuk menduga titik knot dan *bandwidth* pada model regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian.

#### **4.3 Aplikasi pada Data Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan**

Pada bagian ini akan dibahas mengenai pemilihan titik knot dan *bandwidth* yang optimal menggunakan metode CV, GCV dan UBR yang kemudian diaplikasikan pada data persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan pada Tahun 2019.

Kabupaten/Kota yang dijadikan observasi pengamatan pada penelitian ini adalah 53 Kabupaten/Kota yang ada di Pulau Kalimantan. Kota Bontang,

Banjarmasin dan Pontianak tidak digunakan dalam observasi penelitian karena merupakan data pencilan. Kota Bontang merupakan Kota dengan Laju Pertumbuhan Ekonomi yang bernilai negatif. Kota Banjarmasin dan Pontianak merupakan Kota dengan kepadatan penduduk yang tinggi. Sehingga berdasarkan identifikasi tersebut, maka dapat diketahui jika ke 3 Kabupaten/Kota ini merupakan data pencilan dalam variabel prediktor.

### 1. Statistika Deskriptif Variabel Penelitian

Sebelum dilakukan proses pemodelan regresi nonparametrik dengan menggunakan estimator campuran *spline truncated* dan kernel Gaussian pada kasus persentase penduduk miskin di Pulau Kalimantan, maka pada bagian ini akan ditampilkan statistika deskriptif guna memberikan informasi awal dari masing-masing variabel yang digunakan.

**Tabel 4.4** Statistika Deskriptif

Variabel	N	Minimum	Maksimum	Rata-rata	Standar Deviasi
$y$	53	2,42	12,38	6,19	2,39
$x_1$	53	63,24	74,18	70,19	2,44
$x_2$	53	2,15	7,29	4,19	1,39
$x_3$	53	6,00	11,51	8,09	1,17
$x_4$	53	2,52	7,99	5,53	1,16
$x_5$	53	27,21	92,89	68,85	15,78
$x_6$	53	10,26	99,32	51,19	26,29
$t_1$	53	1,36	1297,74	134,81	290,50

Lebih lanjut, statistika deskriptif yang ditampilkan pada Tabel 4.4 memperlihatkan bahwa masing-masing Kabupaten/Kota di pulau Kalimantan memiliki karakteristik yang berbeda-beda untuk semua variabel prediktor maupun respon. Kota Balikpapan merupakan daerah yang memiliki persentase penduduk miskin terendah yaitu sebesar 2,42%, sedangkan Kabupaten Melawi merupakan daerah yang memiliki persentase penduduk miskin tertinggi yaitu sebesar 12,38%, dengan rata-rata persentase penduduk miskin Kabupaten/Kota di pulau Kalimantan adalah sebesar 6,19%. Berikut adalah gambaran karakteristik untuk masing-masing variabel prediktor:

**a. Angka Harapan Hidup ( $X_1$ )**

Rata-rata angka harapan hidup dari Kabupaten/Kota di pulau Kalimantan adalah 70,19 tahun, dengan standar deviasi yaitu 2,44. Kota Balikpapan merupakan daerah dengan angka harapan hidup tertinggi, yaitu sebesar 74,18 tahun, sedangkan Kabupaten Hulu Sungai Utara merupakan daerah dengan angka harapan hidup terendah, yaitu sebesar 63,24 tahun.

**b. Tingkat Pengangguran Terbuka ( $X_2$ )**

Rata-rata tingkat pengangguran terbuka dari Kabupaten/Kota di pulau Kalimantan adalah 4,19%, dengan standar deviasi yaitu 1,39. Kota Balikpapan merupakan daerah dengan tingkat pengangguran terbuka tertinggi, yaitu sebesar 7,29%, sedangkan Kabupaten Pulang Pisau merupakan daerah dengan tingkat pengangguran terbuka terendah, yaitu sebesar 2,15%.

**c. Rata-rata Lama Sekolah ( $X_3$ )**

Rata-rata lama sekolah dari Kabupaten/Kota di pulau Kalimantan adalah 8,09 tahun, dengan standar deviasi yaitu 1,17. Kota Palangka Raya merupakan daerah dengan rata-rata lama sekolah tertinggi, yaitu sebesar 11,51 tahun, sedangkan Kabupaten Kayong Utara merupakan daerah dengan rata-rata lama sekolah terendah, yaitu hanya 6 tahun.

**d. Laju Pertumbuhan Ekonomi ( $X_4$ )**

Rata-rata laju pertumbuhan ekonomi dari Kabupaten/Kota di pulau Kalimantan adalah 5,53%, dengan standar deviasi yaitu 1,16. Kabupaten Ketapang merupakan daerah dengan laju pertumbuhan ekonomi tertinggi, yaitu sebesar 7,99%, sedangkan Kabupaten Penajam Paser Utara merupakan daerah dengan laju pertumbuhan ekonomi terendah, hanya sebesar 2,52%.

**e. Persentase RT dengan Akses Sanitasi Layak ( $X_5$ )**

Rata-rata persentase rumah tangga yang memiliki akses sanitasi layak dari Kabupaten/Kota di pulau Kalimantan adalah 68,85%, dengan standar deviasi yaitu 15,78. Kota Samarinda merupakan daerah dengan persentase rumah tangga yang memiliki akses sanitasi layak tertinggi, yaitu sebesar 92,89%, sedangkan Kabupaten Murung Raya merupakan daerah dengan

persentase rumah tangga yang memiliki akses sanitasi layak terendah, yaitu sebesar 27,21%.

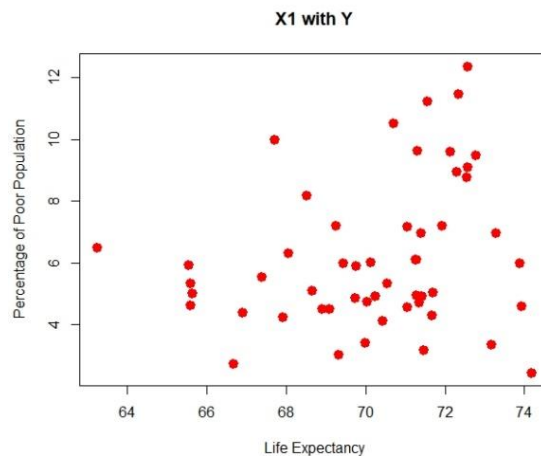
**f. Persentase RT dengan Akses Air Minum Layak ( $X_6$ )**

Rata-rata persentase rumah tangga yang memiliki akses air minum layak dari Kabupaten/Kota di pulau Kalimantan adalah 51,19%, dengan standar deviasi yaitu 26,29. Kota Palangka Raya merupakan daerah dengan persentase rumah tangga yang memiliki akses air minum layak tertinggi, yaitu sebesar 99,32%, sedangkan Kabupaten Mahakam Ulu merupakan daerah dengan persentase rumah tangga yang memiliki akses air minum layak terendah, yaitu hanya sebesar 10,26%.

**g. Kepadatan Penduduk ( $T_1$ )**

Rata-rata kepadatan penduduk dari Kabupaten/Kota di pulau Kalimantan adalah 134,81 jiwa/Km<sup>2</sup>, dengan standar deviasi yaitu 290,50. Kota Balikpapan merupakan daerah dengan penduduk yang padat, dimana kepadatan penduduknya 1297,4 jiwa/Km<sup>2</sup>, sedangkan Kabupaten Mahakam Ulu merupakan daerah dengan kepadatan penduduk yang terendah, yaitu hanya 1,36 jiwa/Km<sup>2</sup>.

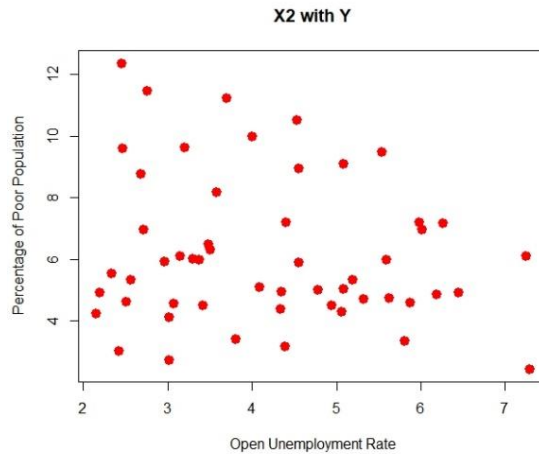
Selanjutnya untuk melihat pola hubungan antara variabel respon dengan masing-masing variabel prediktor dapat dilihat dari diagram pencar. Hasil diagram pencar untuk masing-masing variabel respon dengan variabel prediktor adalah sebagai berikut:



**Gambar 4.3** Diagram pencar antara  $x_1$  dengan  $y$

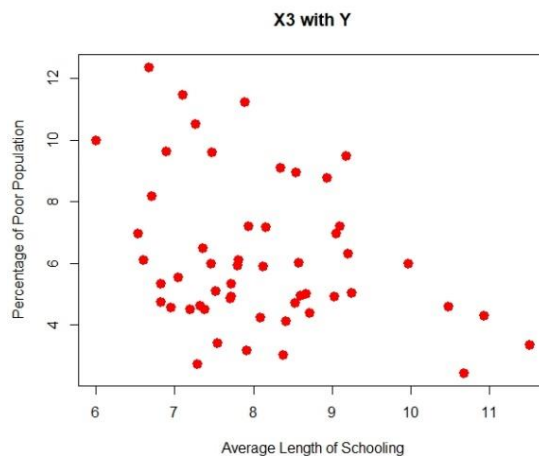


Berdasarkan Gambar 4.3, terlihat pola hubungan antara Angka Harapan Hidup ( $x_1$ ) dengan Persentase Penduduk Miskin ( $y$ ) tidak ada kecenderungan membentuk pola tertentu. Pola hubungan terlihat cenderung mengalami perubahan perilaku pada sub-sub interval tertentu, sehingga variabel Angka Harapan Hidup ( $x_1$ ) akan didekati dengan estimator *spline truncated*.



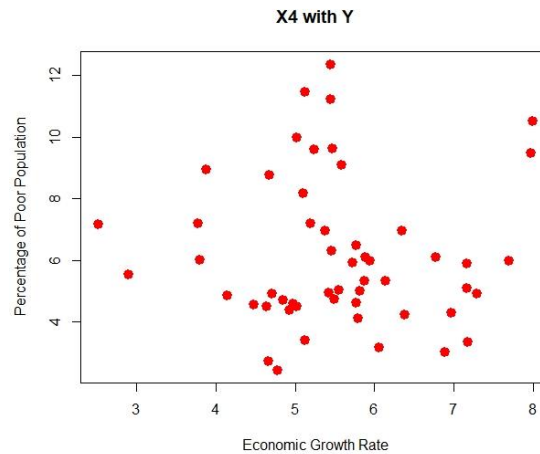
**Gambar 4.4** Diagram pencar antara  $x_2$  dengan  $y$

Berdasarkan Gambar 4.4, terlihat pola hubungan antara Tingkat Pengangguran Terbuka ( $x_2$ ) dengan Persentase Penduduk Miskin ( $y$ ) tidak ada kecenderungan membentuk pola tertentu. Pola hubungan terlihat cenderung mengalami perubahan perilaku pada sub-sub interval tertentu, sehingga variabel Tingkat Pengangguran Terbuka ( $x_2$ ) akan didekati dengan estimator *spline truncated*.



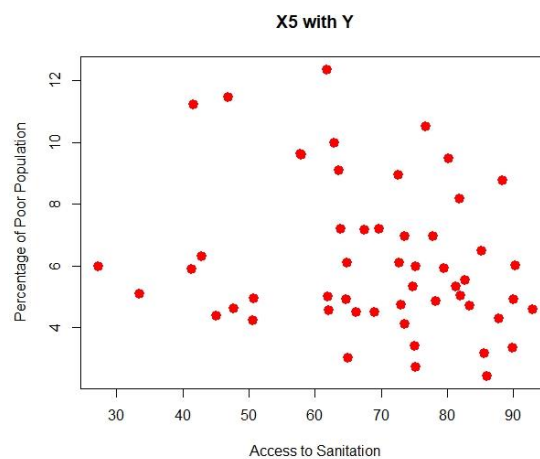
**Gambar 4.5** Diagram pencar antara  $x_3$  dengan  $y$

Berdasarkan Gambar 4.5, terlihat pola hubungan antara Rata-Rata Lama Sekolah ( $x_3$ ) dengan Persentase Penduduk Miskin ( $y$ ) tidak ada kecenderungan membentuk pola tertentu. Pola hubungan terlihat cenderung mengalami perubahan perilaku pada sub-sub interval tertentu, sehingga variabel Rata-Rata Lama Sekolah ( $x_3$ ) akan didekati dengan estimator *spline truncated*.



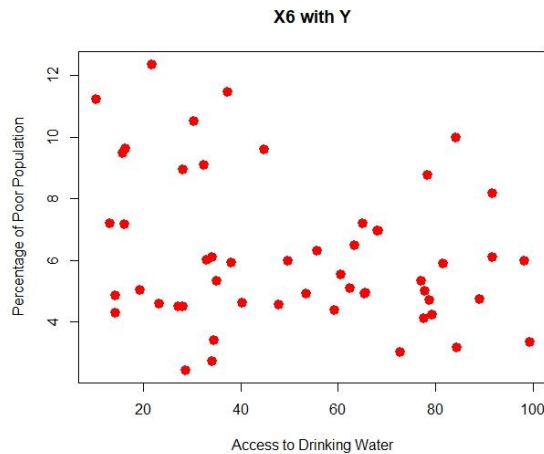
**Gambar 4.6** Diagram pencar antara  $x_4$  dengan  $y$

Berdasarkan Gambar 4.6, terlihat pola hubungan antara Laju Pertumbuhan Ekonomi ( $x_4$ ) dengan Persentase Penduduk Miskin ( $y$ ) tidak ada kecenderungan membentuk pola tertentu. Pola hubungan terlihat cenderung mengalami perubahan perilaku pada sub-sub interval tertentu, sehingga variabel Laju Pertumbuhan Ekonomi ( $x_4$ ) akan didekati dengan estimator *spline truncated*.



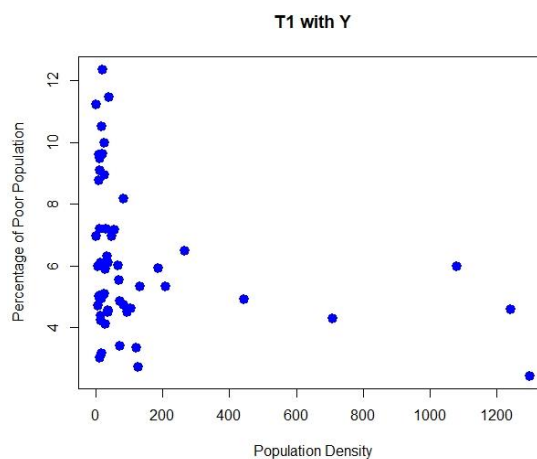
**Gambar 4.7** Diagram pencar antara  $x_5$  dengan  $y$

Berdasarkan Gambar 4.7, terlihat pola hubungan antara Persentase RT dengan Akses Sanitasi Layak ( $x_5$ ) dengan Persentase Penduduk Miskin ( $y$ ) tidak ada kecenderungan membentuk pola tertentu. Pola hubungan terlihat cenderung mengalami perubahan perilaku pada sub-sub interval tertentu, sehingga variabel ( $x_5$ ) akan didekati dengan estimator *spline truncated*.



**Gambar 4.8** Diagram pencar antara  $x_6$  dengan  $y$

Berdasarkan Gambar 4.8, terlihat pola hubungan antara Persentase RT dengan Akses Air Minum Layak ( $x_6$ ) dengan Persentase Penduduk Miskin ( $y$ ) tidak ada kecenderungan membentuk pola tertentu. Pola hubungan terlihat cenderung mengalami perubahan perilaku pada sub-sub interval tertentu, sehingga variabel ( $x_6$ ) akan didekati dengan estimator *spline truncated*.



**Gambar 4.9** Diagram pencar antara  $t_1$  dengan  $y$

Berdasarkan Gambar 4.9, terlihat pola hubungan antara Kepadatan Penduduk ( $t_1$ ) dengan Persentase Penduduk Miskin  $y$  tidak ada kecenderungan membentuk pola tertentu, dan tidak menunjukkan adanya pola yang berubah-ubah pada sub-sub interval tertentu, sehingga dapat dikatakan tidak memiliki suatu pola yang jelas. Maka pola data yang demikian, akan di dekati dengan estimator kernel.

Ringkasan hasil penentuan jenis estimator yang akan digunakan untuk masing-masing variabel prediktor disajikan pada Tabel 4.5.

**Tabel 4.5** Komponen Estimator *Spline Truncated* dan Kernel

Variabel	Simbol Variabel	Deskripsi	Estimator Pemodelan
Prediktor	$x_1$	Angka Harapan Hidup	<i>Spline Truncated</i>
	$x_2$	Tingkat Pengangguran Terbuka	
	$x_3$	Rata-Rata Lama Sekolah	
	$x_4$	Laju Pertumbuhan Ekonomi	
	$x_5$	Persentase RT dengan Akses Sanitasi Layak	
	$x_6$	Persentase RT dengan Akses Air Minum Layak	
	$t_1$	Kepadatan Penduduk	Kernel Gaussian

Model estimator campuran *spline truncated* dan kernel Gaussian yang terbaik, akan di pilih dengan melihat nilai CV, GCV dan UBR terkecil dari beberapa model berdasarkan banyaknya titik knot dan *bandwidth* yang berbeda-beda. Penelitian ini menggunakan jumlah titik knot dengan jumlah yang sama untuk setiap variabel prediktor yang digunakan yaitu 1 sampai dengan 3 titik knot. Kemudian akan ditampilkan 5 posisi titik knot yang berbeda pada masing-masing bagian yang di duga merupakan titik knot dan *bandwidth* yang optimal. Pada penelitian ini akan digunakan 3 metode pemilihan titik knot dan *bandwidth* yang optimal yaitu CV, GCV dan UBR.

## 2. Deteksi Multikolinearitas

Dalam proses pemodelan menggunakan analisis regresi, tidak diperkenankan terdapat masalah atau kasus multikolinearitas. Multikolinearitas itu sendiri merupakan kasus ketika terdapat hubungan yang erat (korelasi yang kuat) antar

variabel prediktor. Pada penelitian ini, akan digunakan nilai *Variance Inflation Factor* (VIF) untuk mendeteksi masalah multikolinearitas.

**Tabel 4.6** Deteksi Multikolinearitas dari Variabel Prediktor

Variabel	Deskripsi	Nilai VIF
$x_1$	Angka Harapan Hidup	1,30
$x_2$	Tingkat Pengangguran Terbuka	1,45
$x_3$	Rata-Rata Lama Sekolah	1,68
$x_4$	Laju Pertumbuhan Ekonomi	1,25
$x_5$	Persentase RT dengan Akses Sanitasi Layak	1,28
$x_6$	Persentase RT dengan Akses Air Minum Layak	1,17
$t_1$	Kepadatan Penduduk	1,57

Berdasarkan Tabel 4.6, maka dapat diketahui jika tidak ada nilai VIF yang melebihi nilai 10, hal tersebut dapat diartikan bahwa tidak terdapat masalah multikolinearitas pada variabel prediktor yang digunakan pada kasus ini.

### 3. Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan menggunakan Metode *Cross-Validation* (CV)

Akan disajikan penerapan pemodelan regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan menggunakan metode *Cross-Validation* (CV) sebagai metode pemilihan titik knot dan *bandwidth* yang optimal.

#### 3.1 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Satu Titik Knot

Pada bagian ini, akan dibahas mengenai pemilihan titik knot dan *bandwidth* yang optimal, diawali dengan menggunakan satu titik knot pada masing-masing variabel prediktor yang di definisikan sebagai komponen *spline truncated*. Nilai CV yang dihasilkan dengan menggunakan satu titik knot pada model regresi nonparametrik estimator campuran ditampilkan pada Tabel 4.7.

**Tabel 4.7** Nilai CV untuk Estimator Campuran dengan Satu Titik Knot

No	Lokasi Titik Knot						Bandwidth	CV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$		
1	69,65	5,16	9,23	5,73	65,71	62,46	2,55	4,27
2	69,27	4,98	9,04	5,54	63,45	59,39	2,60	4,29

**Tabel 4.7** Nilai CV untuk Estimator Campuran dengan Satu Titik Knot (lanjutan)

No	Lokasi Titik Knot						Bandwidth	CV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
3	68,89	4,81	8,85	5,35	61,18	56,32	2,69	4,46
4	70,03	5,34	9,42	5,91	67,97	65,58	2,59	4,51
5	72,67	6,58	10,75	7,23	83,83	87,03	3,01	4,65

Berdasarkan Tabel 4.7, nilai CV minimum adalah sebesar 4,27 dengan *bandwidth* optimal adalah 2,55 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut:

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1 : 69,65$

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1 : 5,16$

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1 : 9,23$

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1 : 5,73$

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1 : 65,71$

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1 : 62,46$

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan satu titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 69,65)_+ + \hat{\phi}_{22}(x_{2i} - 5,16)_+ + \hat{\phi}_{23}(x_{3i} - 9,23)_+ + \hat{\phi}_{24}(x_{4i} - 5,73)_+ + \hat{\phi}_{25}(x_{5i} - 65,71)_+ + \hat{\phi}_{26}(x_{6i} - 62,46)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{2,55} K\left(\frac{t-t_i}{2,55}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{2,55} K\left(\frac{t-t_i}{2,55}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 66,00%.

### 3.2 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Dua Titik Knot

Langkah selanjutnya adalah melakukan pemilihan titik knot dan *bandwidth* yang optimal dengan menggunakan dua titik knot pada masing-masing variabel prediktor yang di definisikan sebagai komponen *spline truncated*. Nilai CV yang dihasilkan dengan menggunakan dua titik knot pada model regresi nonparametrik estimator campuran ditampilkan pada Tabel 4.8.

**Tabel 4.8** Nilai CV untuk Estimator Campuran dengan Dua Titik Knot

No	Lokasi Titik Knot						Bandwidth	CV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	67,01	3,92	7,90	4,40	49,85	40,97	2,18	4,27
	73,42	6,93	11,13	7,61	88,36	93,17		
2	70,40	5,51	9,61	6,10	70,24	68,60	1,96	4,38
	72,29	6,40	10,56	7,04	81,56	83,96		
3	70,03	5,34	9,42	5,91	67,97	65,53	2,02	4,39
	72,29	6,40	10,56	7,04	81,56	83,96		
4	67,38	4,09	8,09	4,59	52,12	44,04	2,18	4,39
	73,42	6,93	11,13	7,61	88,36	93,17		
5	68,89	4,81	8,85	5,35	61,18	56,32	2,11	4,43
	72,29	6,40	10,56	7,04	81,56	83,96		

Berdasarkan Tabel 4.8, nilai CV minimum adalah sebesar 4,27 dengan *bandwidth* optimal adalah 2,18 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut:

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1$  : 67,01       $K_2$  : 73,42

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1$  : 3,92       $K_2$  : 6,93

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1$  : 7,90       $K_2$  : 11,13

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1$  : 4,40       $K_2$  : 7,61

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1$  : 49,85       $K_2$  : 88,36

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1$  : 40,97       $K_2$  : 93,17

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan dua titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 67,01)_+ + \hat{\phi}_{31}(x_{1i} - 73,04)_+ + \hat{\phi}_{22}(x_{2i} - 3,92)_+ + \hat{\phi}_{32}(x_{2i} - 6,93)_+ + \hat{\phi}_{23}(x_{3i} - 7,90)_+ + \hat{\phi}_{33}(x_{3i} - 11,13)_+ + \hat{\phi}_{24}(x_{4i} - 4,40)_+ + \hat{\phi}_{34}(x_{4i} - 7,61)_+ + \hat{\phi}_{25}(x_{5i} - 49,85)_+ + \hat{\phi}_{35}(x_{5i} - 88,36)_+ + \hat{\phi}_{26}(x_{6i} - 40,97)_+ + \hat{\phi}_{36}(x_{6i} - 93,17)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{2,18} K\left(\frac{t-t_i}{2,18}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{2,18} K\left(\frac{t-t_i}{2,18}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 73,35%.

### 3.3 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Tiga Titik Knot

Pada bagian ini akan menampilkan nilai CV yang dihasilkan dengan menggunakan tiga titik knot pada model regresi nonparametrik estimator campuran pada yang ditampilkan pada Tabel 4.9.

**Tabel 4.9** Nilai CV untuk Estimator Campuran dengan Tiga Titik Knot

No	Lokasi Titik Knot						Bandwidth	CV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	65,12	3,03	6,95	3,46	38,53	25,61	1,41	4,19
	66,63	3,74	7,71	4,22	47,59	37,89		
	72,67	6,58	10,75	7,23	83,83	87,03		
2	65,12	3,03	6,95	3,46	38,53	25,61	1,43	4,31
	66,25	3,56	7,52	4,03	45,32	34,82		
	72,67	6,58	10,75	7,23	83,83	87,03		
3	65,50	3,21	7,14	3,65	40,79	28,68	1,43	4,42
	66,63	3,74	7,71	4,21	47,59	37,89		
	72,67	6,58	10,75	7,23	83,83	87,03		
4	65,12	3,03	6,95	3,46	38,53	25,61	1,42	4,62
	67,01	3,92	7,90	4,40	49,85	40,97		
	72,67	6,58	10,75	7,23	83,83	87,03		
5	65,50	3,21	7,14	3,65	40,79	28,68	1,42	4,70
	67,01	3,92	7,90	4,40	49,85	40,97		
	72,67	6,58	10,75	7,23	83,83	87,03		

Berdasarkan Tabel 4.9, nilai CV minimum adalah sebesar 4,19 dengan *bandwidth* optimal adalah 1,41 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut:

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1$  : 65,12       $K_2$  : 66,63       $K_3$  : 72,67

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1$  : 3,03       $K_2$  : 3,74       $K_3$  : 6,58



Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1 : 6,95 \quad K_2 : 7,71 \quad K_3 : 10,75$

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1 : 3,46 \quad K_2 : 4,22 \quad K_3 : 7,23$

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1 : 38,53 \quad K_2 : 47,59 \quad K_3 : 83,83$

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1 : 25,61 \quad K_2 : 37,89 \quad K_3 : 87,03$

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan tiga titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 65,12)_+ + \hat{\phi}_{31}(x_{1i} - 66,63)_+ + \hat{\phi}_{41}(x_{1i} - 72,67)_+ + \hat{\phi}_{22}(x_{2i} - 3,03)_+ + \hat{\phi}_{32}(x_{2i} - 3,74)_+ + \hat{\phi}_{42}(x_{2i} - 6,58)_+ + \hat{\phi}_{23}(x_{3i} - 6,95)_+ + \hat{\phi}_{33}(x_{3i} - 7,71)_+ + \hat{\phi}_{43}(x_{3i} - 10,75)_+ + \hat{\phi}_{24}(x_{4i} - 3,46)_+ + \hat{\phi}_{34}(x_{4i} - 4,22)_+ + \hat{\phi}_{44}(x_{4i} - 7,23)_+ + \hat{\phi}_{25}(x_{5i} - 38,53)_+ + \hat{\phi}_{35}(x_{5i} - 47,59)_+ + \hat{\phi}_{45}(x_{5i} - 83,83)_+ + \hat{\phi}_{26}(x_{6i} - 25,61)_+ + \hat{\phi}_{36}(x_{6i} - 37,89)_+ + \hat{\phi}_{46}(x_{6i} - 87,03)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,41} K\left(\frac{t-t_i}{1,41}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,41} K\left(\frac{t-t_i}{1,41}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 83,80%.

### 3.4 Ringkasan Hasil Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Metode *Cross-Validation* (CV)

Setelah dilakukan pemodelan dengan menggunakan pendekatan regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian dengan metode *Cross-Validation* (CV) sebagai metode pemilihan titik knot dan *bandwidth* yang optimal, maka akan ditampilkan ringkasan seperti pada Tabel 4.10.

**Tabel 4.10** Ringkasan dengan Metode *Cross-Validation* (CV)

Titik Knot	Nilai CV	Koefisien Determinasi ( $R^2$ )
Satu titik knot	4,27	66,00%
Dua titik knot	4,27	73,35%
<b>Tiga titik knot</b>	<b>4,19</b>	<b>83,80%</b>

Berdasarkan Tabel 4.10, terlihat bahwa model regresi nonparametrik dengan estimator campuran *spline truncated* dan kernel Gaussian yang menghasilkan nilai CV minimum dan Koefisien Determinasi ( $R^2$ ) terbesar adalah model regresi nonparametrik estimator campuran dengan tiga titik knot.

#### 4. Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan menggunakan Metode *Generalized Cross-Validation* (GCV)

Akan disajikan penerapan pemodelan regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan menggunakan metode *Generalized Cross-Validation* (GCV) sebagai metode pemilihan titik knot dan *bandwidth* yang optimal.

##### 4.1 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Satu Titik Knot

Pada bagian ini, akan dibahas mengenai pemilihan titik knot dan *bandwidth* yang optimal, diawali dengan menggunakan satu titik knot pada masing-masing variabel prediktor yang di definisikan sebagai komponen *spline truncated*. Nilai GCV yang dihasilkan dengan menggunakan satu titik knot pada model regresi nonparametrik estimator campuran ditampilkan pada Tabel 4.11.

**Tabel 4.11** Nilai GCV untuk Estimator Campuran dengan Satu Titik Knot

No	Lokasi Titik Knot						Bandwidth	GCV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	69,65	5,16	9,23	5,73	65,71	62,47	2,55	5,88
2	70,03	5,34	9,42	5,91	67,97	65,53	2,59	5,99
3	69,27	4,98	9,04	5,53	63,44	59,39	2,60	6,00
4	68,89	4,80	8,85	5,34	61,18	56,32	2,69	6,23
5	70,40	5,51	9,61	6,10	70,24	68,60	2,71	6,27

Berdasarkan Tabel 4.11, nilai GCV minimum adalah sebesar 5,88 dengan *bandwidth* optimal adalah 2,55 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut:

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1 : 69,65$

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1 : 5,16$

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1 : 9,23$

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1 : 5,73$

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1 : 65,71$

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1 : 62,47$

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan satu titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_1 x_{1i} + \hat{\delta}_2 x_{2i} + \hat{\delta}_3 x_{3i} + \hat{\delta}_4 x_{4i} + \hat{\delta}_5 x_{5i} + \hat{\delta}_6 x_{6i} + \hat{\phi}_{21}(x_{1i} - 69, 65)_+ + \hat{\phi}_{22}(x_{2i} - 5, 16)_+ + \hat{\phi}_{23}(x_{3i} - 9, 23)_+ + \hat{\phi}_{24}(x_{4i} - 5, 72)_+ + \hat{\phi}_{25}(x_{5i} - 65, 71)_+ + \hat{\phi}_{26}(x_{6i} - 62, 47)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{2,55} K\left(\frac{t-t_i}{2,55}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{2,55} K\left(\frac{t-t_i}{2,55}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 66,00%.

#### 4.2 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Dua Titik Knot

Langkah selanjutnya adalah melakukan pemilihan titik knot dan *bandwidth* yang optimal dengan menggunakan dua titik knot pada masing-masing variabel prediktor yang di definisikan sebagai komponen *spline truncated*. Berikut akan disajikan nilai GCV yang dihasilkan dengan menggunakan dua titik knot pada Tabel 4.12.

**Tabel 4.12** Nilai GCV untuk Estimator Campuran dengan Dua Titik Knot

No	Lokasi Titik Knot						Bandwidth	GCV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	71,16	5,87	9,99	6,48	74,77	74,75	1,82	3,09
	72,29	6,40	10,56	7,05	81,56	83,96		
2	70,78	5,69	9,80	6,29	72,50	71,68	1,85	3,14
	72,29	6,40	10,56	7,05	81,56	83,96		
3	70,78	5,69	9,80	6,29	72,50	71,68	1,89	3,20
	72,67	6,58	10,75	7,23	83,83	87,03		
4	70,41	5,52	9,61	6,10	70,24	68,60	1,92	3,26
	72,67	6,58	10,75	7,23	83,83	87,03		
5	71,16	5,87	9,99	6,48	74,77	74,75	1,92	3,26
	72,67	6,58	10,75	7,23	83,83	87,03		

Berdasarkan Tabel 4.12, nilai GCV minimum adalah sebesar 3,09 dengan *bandwidth* optimal adalah 1,82 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1 : 71,16 \quad K_2 : 72,29$

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1 : 5,87 \quad K_2 : 6,40$

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1 : 9,99 \quad K_2 : 10,56$

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1 : 6,48 \quad K_2 : 7,05$

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1 : 74,77 \quad K_2 : 81,56$

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1 : 74,75 \quad K_2 : 83,96$

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan dua titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 71,16)_+ + \hat{\phi}_{31}(x_{1i} - 72,29)_+ + \hat{\phi}_{22}(x_{2i} - 5,87)_+ + \hat{\phi}_{32}(x_{2i} - 6,40)_+ + \hat{\phi}_{23}(x_{3i} - 9,99)_+ + \hat{\phi}_{33}(x_{3i} - 10,56)_+ + \hat{\phi}_{24}(x_{4i} - 6,48)_+ + \hat{\phi}_{34}(x_{4i} - 7,05)_+ + \hat{\phi}_{25}(x_{5i} - 74,77)_+ + \hat{\phi}_{35}(x_{5i} - 81,56)_+ + \hat{\phi}_{26}(x_{6i} - 74,75)_+ + \hat{\phi}_{36}(x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,82} K\left(\frac{t-t_i}{1,82}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,82} K\left(\frac{t-t_i}{1,82}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 85,10%.

### 4.3 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Tiga Titik Knot

Bagian ini akan menampilkan nilai GCV yang dihasilkan dengan menggunakan tiga titik knot pada model regresi nonparametrik estimator campuran yang selanjutnya ditampilkan pada Tabel 4.13.

**Tabel 4.13** Nilai GCV untuk Estimator Campuran dengan Tiga Titik Knot

No	Lokasi Titik Knot						Bandwidth	GCV
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	64,37	2,68	6,57	3,08	34,00	19,47	1,09	1,51
	71,54	6,05	10,18	6,66	77,03	77,82		
	72,29	6,40	10,56	7,05	81,56	83,96		
2	64,74	2,85	6,76	3,27	36,26	22,54	1,09	1,52
	71,54	6,04	10,18	6,66	77,03	77,82		
	72,29	6,40	10,56	7,04	81,56	83,96		
3	65,12	3,03	6,95	3,46	38,53	25,61	1,10	1,54
	71,53	6,04	10,18	6,66	77,03	77,82		
	72,29	6,40	10,56	7,04	81,56	83,96		
4	63,99	2,50	6,38	2,89	31,73	16,40	1,11	1,54
	71,54	6,05	10,18	6,66	77,03	77,82		
	72,29	6,40	10,56	7,04	81,56	83,96		
5	65,12	3,03	6,95	3,46	38,53	25,61	1,14	1,58
	71,16	5,87	9,99	6,48	74,77	74,75		
	72,29	6,40	10,56	7,04	81,56	83,96		

Berdasarkan Tabel 4.13, nilai GCV minimum adalah sebesar 1,51 dengan *bandwidth* optimal adalah 1,09 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1$  : 64,37       $K_2$  : 71,54       $K_3$  : 72,30

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1$  : 2,68       $K_2$  : 6,05       $K_3$  : 6,40

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1$  : 6,57       $K_2$  : 10,18       $K_3$  : 10,56

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1$  : 3,08       $K_2$  : 6,66       $K_3$  : 7,05

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1$  : 34,00       $K_2$  : 77,03       $K_3$  : 81,56

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1$  : 19,47       $K_2$  : 77,82       $K_3$  : 83,96

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan tiga titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 64, 37)_+ + \hat{\phi}_{31}(x_{1i} - 71, 54)_+ + \hat{\phi}_{41}(x_{1i} - 72, 30)_+ + \hat{\phi}_{22}(x_{2i} - 2, 68)_+ + \hat{\phi}_{32}(x_{2i} - 6, 05)_+ + \hat{\phi}_{42}(x_{2i} - 6, 40)_+ + \hat{\phi}_{23}(x_{3i} - 6, 57)_+ + \hat{\phi}_{33}(x_{3i} - 10, 18)_+ + \hat{\phi}_{43}(x_{3i} - 10, 56)_+ + \hat{\phi}_{24}(x_{4i} - 3, 08)_+ + \hat{\phi}_{34}(x_{4i} - 6, 66)_+ + \hat{\phi}_{44}(x_{4i} - 7, 05)_+ + \hat{\phi}_{25}(x_{5i} - 34, 00)_+ + \hat{\phi}_{35}(x_{5i} - 77, 03)_+ + \hat{\phi}_{45}(x_{5i} - 81, 56)_+ + \hat{\phi}_{26}(x_{6i} - 19, 47)_+ + \hat{\phi}_{36}(x_{6i} - 77, 82)_+ + \hat{\phi}_{46}(x_{6i} - 83, 96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 90,52%.

#### 4.4 Ringkasan Hasil Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Metode *Generalized Cross-Validation* (GCV)

Setelah dilakukan pemodelan dengan menggunakan pendekatan regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian dengan metode *Generalized Cross-Validation* (GCV) sebagai metode pemilihan titik knot dan *bandwidth* yang optimal, maka akan ditampilkan ringkasan seperti pada Tabel 4.14.

**Tabel 4.14** Ringkasan dengan Metode *Generalized Cross-Validation* (GCV)

Titik Knot	Nilai GCV	Koefisien Determinasi ( $R^2$ )
Satu titik knot	5,88	66,00%
Dua titik knot	3,09	85,10%
<b>Tiga titik knot</b>	<b>1,51</b>	<b>90,52%</b>

Berdasarkan Tabel 4.14, terlihat bahwa model regresi nonparametrik dengan estimator campuran *spline truncated* dan kernel Gaussian yang menghasilkan nilai GCV minimum serta Koefisien Determinasi ( $R^2$ ) terbesar adalah model regresi nonparametrik dengan tiga titik knot.

## 5. Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Pulau Kalimantan menggunakan Metode *Unbiased Risk* (UBR)

Akan disajikan penerapan pemodelan regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan menggunakan

kriteria *Unbiased Risk* (UBR) sebagai metode pemilihan titik knot dan *bandwidth* yang optimal.

### 5.1 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Satu Titik Knot

Pada bagian ini, akan dibahas mengenai pemilihan titik knot dan *bandwidth* yang optimal, diawali dengan menggunakan satu titik knot pada masing-masing variabel prediktor yang di definisikan sebagai komponen *spline truncated*. Nilai UBR yang dihasilkan dengan menggunakan satu titik knot pada model regresi nonparametrik estimator campuran ditampilkan pada Tabel 4.15.

**Tabel 4.15** Nilai UBR untuk Estimator Campuran dengan Satu Titik Knot

No	Lokasi Titik Knot						Bandwidth	UBR
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	73,80	7,11	11,32	7,80	90,62	96,24	3,09	$2,68 \times 10^{-5}$
2	73,42	6,93	11,13	7,61	88,36	93,17	3,12	$4,56 \times 10^{-5}$
3	73,04	6,75	10,94	7,42	86,09	90,10	3,04	$2,10 \times 10^{-4}$
4	72,67	6,58	10,75	7,23	83,83	87,03	3,01	$2,93 \times 10^{-4}$
5	72,29	6,40	10,56	7,04	81,56	83,96	3,15	$1,31 \times 10^{-3}$

Berdasarkan Tabel 4.15, nilai UBR minimum adalah sebesar  $2,68 \times 10^{-5}$  dengan *bandwidth* optimal adalah 3,09 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut:

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1$  : 73,80

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1$  : 7,11

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1$  : 11,32

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1$  : 7,80

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1$  : 90,62

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1$  : 96,24

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan satu titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 73, 80)_+ + \hat{\phi}_{22}(x_{2i} - 7, 11)_+ + \hat{\phi}_{23}(x_{3i} - 11, 32)_+ + \hat{\phi}_{24}(x_{4i} - 7, 80)_+ + \hat{\phi}_{25}(x_{5i} - 90, 62)_+ + \hat{\phi}_{26}(x_{6i} - 96, 24)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{3,09} K\left(\frac{t-t_i}{3,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{3,09} K\left(\frac{t-t_i}{3,09}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 60,13%.

## 5.2 Pemilihan Titik Knot dan *Bandwidth* Optimal dengan Dua Titik Knot

Langkah selanjutnya adalah melakukan pemilihan titik knot dan *bandwidth* yang optimal dengan menggunakan dua titik knot pada masing-masing variabel prediktor yang di definisikan sebagai komponen *spline truncated*. Berikut akan disajikan nilai UBR yang dihasilkan dengan menggunakan dua titik knot pada Tabel 4.16.

**Tabel 4.16** Nilai UBR untuk Estimator Campuran dengan Dua Titik Knot

No	Lokasi Titik Knot						Bandwidth	UBR
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	71,92	6,22	10,37	6,86	79,30	80,89	2,58	$1,01 \times 10^{-5}$
	73,80	7,11	11,32	7,80	90,62	96,24		
2	69,27	4,98	9,04	5,53	63,44	59,39	2,21	$1,33 \times 10^{-5}$
	73,80	7,11	11,32	7,80	90,62	96,24		
3	63,61	2,32	6,19	2,70	29,47	13,33	2,59	$1,54 \times 10^{-5}$
	73,42	6,93	11,13	7,61	88,36	93,17		
4	70,03	5,34	9,42	5,92	67,97	65,53	2,38	$1,95 \times 10^{-5}$
	70,78	5,69	9,80	6,29	72,50	71,68		
5	71,92	6,22	10,37	6,85	79,30	80,89	2,57	$2,26 \times 10^{-5}$
	73,42	6,93	11,13	7,61	88,36	93,17		

Berdasarkan Tabel 4.16, nilai UBR minimum adalah sebesar  $1,01 \times 10^{-5}$  dengan *bandwidth* optimal adalah 2,58 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut

Pada variabel  $X_1$  (Angka Harapan Hidup)

$K_1$  : 71,92       $K_2$  : 73,80

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$K_1$  : 6,22       $K_2$  : 7,11



Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$K_1 : 10,37 \quad K_2 : 11,32$

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$K_1 : 6,86 \quad K_2 : 7,80$

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$K_1 : 79,30 \quad K_2 : 90,62$

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$K_1 : 80,89 \quad K_2 : 96,24$

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan dua titik knot yang terbaik.

$$\hat{y} = \hat{\delta}_0^* + \hat{\delta}_{11}x_{1i} + \hat{\delta}_{12}x_{2i} + \hat{\delta}_{13}x_{3i} + \hat{\delta}_{14}x_{4i} + \hat{\delta}_{15}x_{5i} + \hat{\delta}_{16}x_{6i} + \hat{\phi}_{21}(x_{1i} - 71, 92)_+ + \hat{\phi}_{31}(x_{1i} - 73, 80)_+ + \hat{\phi}_{22}(x_{2i} - 6, 22)_+ + \hat{\phi}_{32}(x_{2i} - 7, 11)_+ + \hat{\phi}_{23}(x_{3i} - 10, 37)_+ + \hat{\phi}_{33}(x_{3i} - 11, 32)_+ + \hat{\phi}_{24}(x_{4i} - 6, 86)_+ + \hat{\phi}_{34}(x_{4i} - 7, 80)_+ + \hat{\phi}_{25}(x_{5i} - 79, 30)_+ + \hat{\phi}_{35}(x_{5i} - 90, 62)_+ + \hat{\phi}_{26}(x_{6i} - 80, 89)_+ + \hat{\phi}_{36}(x_{6i} - 96, 24)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{2,58} K\left(\frac{t-t_i}{2,58}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{2,58} K\left(\frac{t-t_i}{2,58}\right)} \right] y_i$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 70,12%.

### 5.3 Pemilihan Titik Knot dan *Bandwidth* Optimal Tiga Titik Knot

Bagian ini akan menampilkan nilai UBR yang dihasilkan dengan menggunakan tiga titik knot pada model regresi nonparametrik estimator campuran yang ditampilkan pada Tabel 4.17.

**Tabel 4.17** Nilai UBR untuk Estimator Campuran dengan Tiga Titik Knot

No	Lokasi Titik Knot						Bandwidth	UBR
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
1	64,37	2,68	6,57	3,08	34,00	19,47	1,92	$1,79 \times 10^{-6}$
	71,92	6,22	10,37	6,86	79,30	80,89		
	73,80	7,11	11,32	7,80	90,62	96,24		
2	63,61	2,33	6,19	2,71	29,47	13,33	1,88	$2,17 \times 10^{-6}$
	70,40	5,52	9,61	6,10	70,24	68,61		
	73,80	7,11	11,32	7,80	90,62	96,24		
3	65,12	3,03	6,95	3,46	38,53	25,61	1,96	$2,27 \times 10^{-6}$
	71,92	6,23	10,37	6,85	79,30	80,89		
	73,80	7,11	11,32	7,80	90,63	96,24		
4	70,40	5,52	9,61	6,10	70,24	68,60	1,94	$2,66 \times 10^{-6}$
	71,54	6,05	10,18	6,66	77,03	77,82		
	73,80	7,11	11,32	7,80	90,62	96,24		

**Tabel 4.17** Nilai UBR untuk Estimator Campuran dengan Tiga Titik Knot (lanjutan)

No	Lokasi Titik Knot						Bandwidth	UBR
	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$	
5	70,40	5,51	9,61	6,10	70,24	68,60	1,86	$2,75 \times 10^{-6}$
	71,91	6,22	10,37	6,85	79,30	80,89		
	73,42	6,93	11,13	7,61	88,36	93,17		

Berdasarkan Tabel 4.17, nilai UBR minimum adalah sebesar  $1,79 \times 10^{-6}$  dengan *bandwidth* optimal adalah 1,92 serta lokasi titik knot optimal untuk setiap variabel prediktor komponen *spline truncated* adalah sebagai berikut

Pada variabel  $X_1$  (Angka Harapan Hidup)

$$K_1 : 64,37 \quad K_2 : 71,91 \quad K_3 : 73,80$$

Pada variabel  $X_2$  (Tingkat Pengangguran Terbuka)

$$K_1 : 2,68 \quad K_2 : 6,22 \quad K_3 : 7,11$$

Pada variabel  $X_3$  (Rata-Rata Lama Sekolah)

$$K_1 : 6,57 \quad K_2 : 10,37 \quad K_3 : 11,32$$

Pada variabel  $X_4$  (Laju Pertumbuhan Ekonomi)

$$K_1 : 3,08 \quad K_2 : 6,85 \quad K_3 : 7,80$$

Pada variabel  $X_5$  (Persentase RT dengan Akses Sanitasi Layak)

$$K_1 : 34,00 \quad K_2 : 79,30 \quad K_3 : 90,62$$

Pada variabel  $X_6$  (Persentase RT dengan Akses Air Minum Layak)

$$K_1 : 19,47 \quad K_2 : 80,89 \quad K_3 : 96,24$$

Selanjutnya disajikan model regresi nonparametrik menggunakan estimator campuran *spline truncated* dan kernel Gaussian dengan tiga titik knot yang terbaik.

$$\begin{aligned} \hat{y} = & \hat{\delta}_0^* + \hat{\delta}_{11} x_{1i} + \hat{\delta}_{12} x_{2i} + \hat{\delta}_{13} x_{3i} + \hat{\delta}_{14} x_{4i} + \hat{\delta}_{15} x_{5i} + \hat{\delta}_{16} x_{6i} + \hat{\phi}_{21} (x_{1i} - 64,37)_+ + \hat{\phi}_{31} (x_{1i} - 71,91)_+ + \\ & \hat{\phi}_{41} (x_{1i} - 73,80)_+ + \hat{\phi}_{22} (x_{2i} - 2,68)_+ + \hat{\phi}_{32} (x_{2i} - 6,22)_+ + \hat{\phi}_{42} (x_{2i} - 7,11)_+ + \hat{\phi}_{23} (x_{3i} - 6,57)_+ + \\ & \hat{\phi}_{33} (x_{3i} - 10,37)_+ + \hat{\phi}_{43} (x_{3i} - 11,32)_+ + \hat{\phi}_{24} (x_{4i} - 3,08)_+ + \hat{\phi}_{34} (x_{4i} - 6,85)_+ + \hat{\phi}_{44} (x_{4i} - 7,80)_+ + \\ & \hat{\phi}_{25} (x_{5i} - 34,00)_+ + \hat{\phi}_{35} (x_{5i} - 79,30)_+ + \hat{\phi}_{45} (x_{5i} - 90,62)_+ + \hat{\phi}_{26} (x_{6i} - 19,47)_+ + \hat{\phi}_{36} (x_{6i} - 80,89)_+ + \\ & \hat{\phi}_{46} (x_{6i} - 96,24)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,92} K\left(\frac{t-t_i}{1,92}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,92} K\left(\frac{t-t_i}{1,92}\right)} \right] y_i \end{aligned}$$

Didapatkan nilai koefisien determinasi ( $R^2$ ) sebesar 79,24%.

#### 5.4 Ringkasan Hasil Pemilihan Titik Knot dan *Bandwidth* Optimal

Setelah dilakukan pemodelan dengan menggunakan pendekatan regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian dengan metode *Unbiased Risk* (UBR) sebagai metode pemilihan titik knot dan *bandwidth* yang optimal, maka akan ditampilkan ringkasan seperti pada Tabel 4.18.

**Tabel 4.18** Ringkasan dengan Metode *Unbiased Risk* (UBR)

Titik Knot	Nilai UBR	Koefisien Determinasi ( $R^2$ )
Satu titik knot	$2,68 \times 10^{-5}$	60,13%
Dua titik knot	$1,01 \times 10^{-5}$	70,12%
<b>Tiga titik knot</b>	<b><math>1,79 \times 10^{-6}</math></b>	<b>79,24%</b>

Berdasarkan Tabel 4.18, terlihat bahwa model regresi nonparametrik dengan estimator campuran *spline truncated* dan kernel Gaussian yang menghasilkan nilai UBR minimum serta Koefisien Determinasi ( $R^2$ ) terbesar adalah model regresi nonparametrik dengan tiga titik knot.

#### 6. Ringkasan Hasil Metode CV, GCV dan UBR dalam Pemodelan Regresi Nonparametrik menggunakan Estimator Campuran

Pada bagian ini akan ditampilkan hasil perbandingan metode pemilihan titik knot dan *bandwidth* yang terbaik menggunakan metode *Cross-Validation* (CV), *Generalized Cross-Validation* (GCV) dan *Unbiased Risk* (UBR) pada pemodelan data persentase penduduk miskin Kabupaten/Kota di pulau Kalimantan.

**Tabel 4.19** Perbandingan Metode CV, GCV dan UBR

	CV	GCV	UBR
<b>Jumlah Titik Knot</b>	3	3	3
<b>Lokasi Titik Knot</b>	$X_1 = [K_1: 65,12; K_2: 66,63; K_3: 72,67]$ $X_2 = [K_1: 3,03; K_2: 3,74; K_3: 6,58]$ $X_3 = [K_1: 6,95; K_2: 7,71; K_3: 10,75]$ $X_4 = [K_1: 3,46; K_2: 4,22; K_3: 7,23]$ $X_5 = [K_1: 38,53; K_2: 47,59; K_3: 83,83]$ $X_6 = [K_1: 25,61; K_2: 37,89; K_3: 87,03]$	$X_1 = [K_1: 64,37; K_2: 71,54; K_3: 72,30]$ $X_2 = [K_1: 2,68; K_2: 6,05; K_3: 6,40]$ $X_3 = [K_1: 6,57; K_2: 10,18; K_3: 10,56]$ $X_4 = [K_1: 3,08; K_2: 6,66; K_3: 7,05]$ $X_5 = [K_1: 34,00; K_2: 77,03; K_3: 81,56]$ $X_6 = [K_1: 19,47; K_2: 77,82; K_3: 83,96]$	$X_1 = [K_1: 64,37; K_2: 71,92; K_3: 73,80]$ $X_2 = [K_1: 2,68; K_2: 6,22; K_3: 7,11]$ $X_3 = [K_1: 6,57; K_2: 10,37; K_3: 11,32]$ $X_4 = [K_1: 3,08; K_2: 6,86; K_3: 7,80]$ $X_5 = [K_1: 34,00; K_2: 79,30; K_3: 90,62]$ $X_6 = [K_1: 19,47; K_2: 80,89; K_3: 96,24]$

**Tabel 4.19** Perbandingan Metode CV, GCV dan UBR (lanjutan)

	CV	GCV	UBR
<b>Bandwidth</b>	1,41	1,09	1,92
<b>Koefisien Determinasi (<math>R^2</math>)</b>	83,80%	90,52%	79,24%

Berdasarkan Tabel 4.19, dapat diketahui bahwa dengan menggunakan metode GCV dalam proses pemilihan titik knot dan *bandwidth* yang optimal pada model regresi nonparametrik estimator campuran, diperoleh hasil yang lebih baik dibandingkan dengan metode CV maupun UBR. Hal ini dikarenakan, dengan metode GCV didapatkan performa dan akurasi dari model regresi nonparametrik estimator campuran yang lebih baik, dengan nilai koefisien determinasi ( $R^2$ ) sebesar 90,52%. Selanjutnya dapat pula diketahui, jika lokasi titik knot yang didapatkan dari setiap metode yang dicobakan berbeda-beda.

Kemudian tahap selanjutnya dilakukan pemeriksaan dengan melakukan pengujian asumsi residual distribusi normal dari model regresi nonparametrik estimator campuran yang terbaik menggunakan statistik uji Kolmogorov-Smirnov. Hipotesis yang digunakan pada pengujian distribusi Normal adalah sebagai berikut:

$H_0$  : Residual berdistribusi Normal

$H_1$  : Residual tidak berdistribusi Normal

Dengan taraf signifikansi 5% maka:

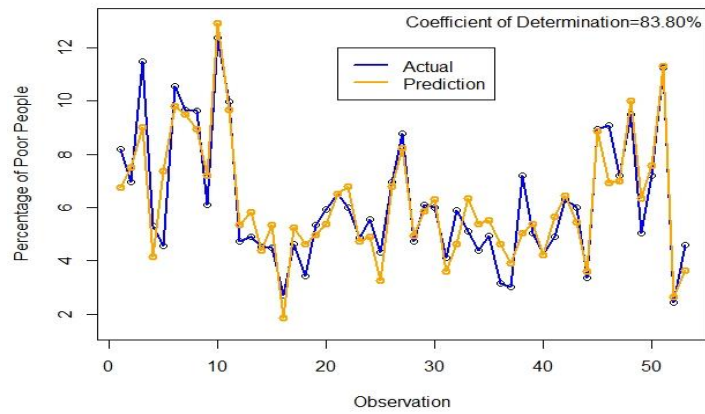
**Tabel 4.20** Uji Normalitas dengan Kolmogorov-Smirnov Test

	P-Value	Keputusan
Model Regresi Nonparametrik Estimator Campuran Terbaik dengan Metode GCV	0,402	Gagal Tolak $H_0$

Berdasarkan Tabel 4.20, diketahui jika residual yang dihasilkan model regresi nonparametrik estimator campuran mengikuti distribusi Normal, dengan nilai P-Value yang didapat lebih besar dari 0,05, sehingga diputuskan Gagal Tolak  $H_0$ .

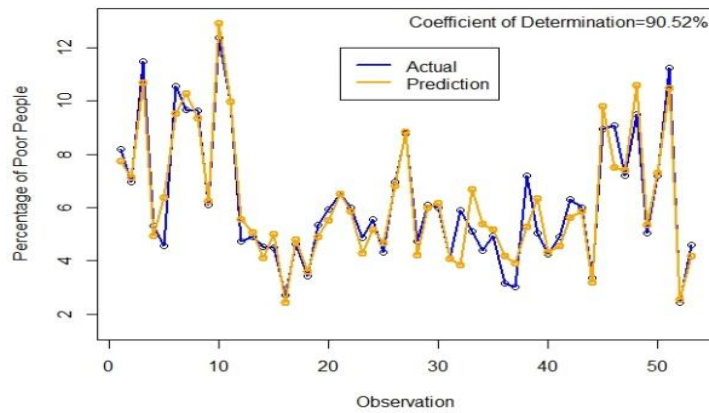
Pada kasus ini pula, didapatkan hasil jika metode GCV lebih baik dalam menduga titik knot dan *bandwidth* yang optimal sedemikian sehingga diperoleh nilai Koefisien Determinasi ( $R^2$ ) yang tertinggi dibandingkan kedua metode lainnya, yaitu CV dan UBR. Dapat dilihat kembali pada Tabel 4. 19.

**Mixed Estimator Spline & Kernel Gaussian with CV Method**



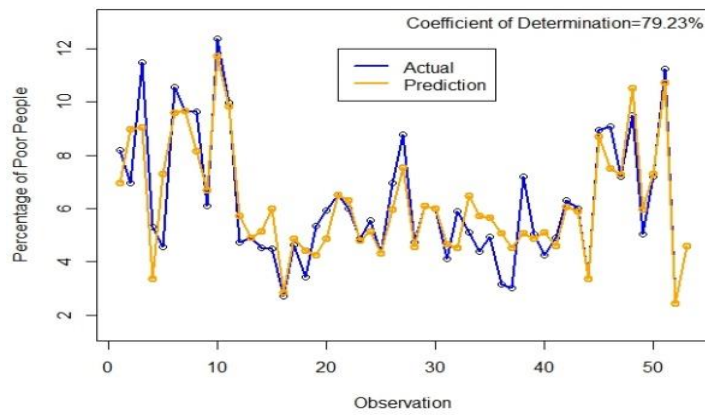
**Gambar 4.10** Visualisasi nilai  $y$  dengan  $\hat{y}$  dengan metode CV

**Mixed Estimator Spline & Kernel Gaussian with GCV Method**



**Gambar 4.11** Visualisasi nilai  $y$  dengan  $\hat{y}$  dengan metode GCV

**Mixed Estimator Spline & Kernel Gaussian with UBR Method**



**Gambar 4.12** Visualisasi nilai  $y$  dengan  $\hat{y}$  dengan metode UBR

Gambar 4.10, 4.11 dan 4.12 merupakan visualisasi nilai  $y$  dengan  $\hat{y}$  dari setiap model yang terbaik berdasarkan metode CV, GCV dan UBR.

## 7. Estimasi Parameter dan Interpretasi Model

Pada bagian ini, akan ditampilkan hasil estimasi parameter regresi nonparametrik dengan estimator campuran menggunakan metode GCV:

**Tabel 4.21** Estimasi Parameter Model Regresi Nonparametrik

Variabel	Estimasi Parameter
Konstanta	$\hat{\delta}_0^* = -6,14 \times 10^{-1}$
$x_1 =$ Angka Harapan Hidup	$\hat{\delta}_{11} = 5,66 \times 10^{-1}$
	$\hat{\phi}_{21} = -6,53 \times 10^{-1}$
	$\hat{\phi}_{31} = 4,05$
	$\hat{\phi}_{41} = -6,99$
$x_2 =$ Tingkat Pengangguran Terbuka	$\hat{\delta}_{12} = 1,54$
	$\hat{\phi}_{22} = -1,64$
	$\hat{\phi}_{32} = 1,41 \times 10^{-1}$
	$\hat{\phi}_{42} = 1,64$
$x_3 =$ Rata-Rata Lama Sekolah	$\hat{\delta}_{13} = -4,69$
	$\hat{\phi}_{23} = 4,05$
	$\hat{\phi}_{33} = 1,09 \times 10^1$
	$\hat{\phi}_{43} = -1,49 \times 10^1$
$x_4 =$ Laju Pertumbuhan Ekonomi	$\hat{\delta}_{14} = -1,32$
	$\hat{\phi}_{24} = 1,20$
	$\hat{\phi}_{34} = -3,82$
	$\hat{\phi}_{44} = 1,03 \times 10^1$
$x_5 =$ Persentase RT dengan Akses Sanitasi Layak	$\hat{\delta}_{15} = -6,25 \times 10^{-4}$
	$\hat{\phi}_{25} = -2,49 \times 10^{-2}$
	$\hat{\phi}_{35} = 1,22 \times 10^{-1}$
	$\hat{\phi}_{45} = 3,34 \times 10^{-3}$
$x_6 =$ Persentase RT dengan Akses Air Minum Layak	$\hat{\delta}_{16} = -1,76 \times 10^{-1}$
	$\hat{\phi}_{26} = 1,83 \times 10^{-1}$
	$\hat{\phi}_{36} = -2,58 \times 10^{-1}$
	$\hat{\phi}_{46} = 5,02 \times 10^{-1}$
<i>Bandwidth</i>	$\alpha = 1,09$

Selanjutnya hasil estimasi parameter pada Tabel 4.21 akan membentuk persamaan model regresi nonparametrik estimator campuran fungsi *spline truncated* linier multivariabel dan kernel Gaussian dengan tiga titik knot sebagai berikut:

$$\hat{y} = -6,13 \times 10^{-1} + 5,66 \times 10^{-1} x_{1i} + 1,54 x_{2i} - 4,69 x_{3i} - 1,32 x_{4i} - 6,25 \times 10^{-4} x_{5i} - 1,76 \times 10^{-1} x_{6i} - 6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ - 1,64 (x_{2i} - 2,68)_+ + 1,41 \times 10^{-1} (x_{2i} - 6,05)_+ + 1,64 (x_{2i} - 6,40)_+ + 4,05 (x_{3i} - 6,57)_+ + 1,09 \times 10^1 (x_{3i} - 10,18)_+ - 1,49 \times 10^1 (x_{3i} - 10,56)_+ + 1,20 (x_{4i} - 3,08)_+ - 3,82 (x_{4i} - 6,66)_+ + 1,03 \times 10^1 (x_{4i} - 7,05)_+ - 2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)} \right] y_i$$

Akan dilakukan interpretasi model regresi nonparametrik dengan estimator campuran fungsi *spline truncated* dan kernel Gaussian. Untuk kasus ini, model pola data yang di definisikan menjadi komponen *spline truncated* dapat diinterpretasikan, namun untuk pola data yang didefinisikan menjadi komponen Kernel Gaussian tidak dapat diinterpretasikan.

**a) Model untuk variabel Angka Harapan Hidup (X<sub>1</sub>)**

Akan diasumsikan data selain angka harapan hidup (X<sub>1</sub>) di notasikan sebagai c<sub>1</sub>, sehingga diperoleh model umumnya:

$$\hat{y} = 5,66 \times 10^{-1} x_{1i} - 6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ + c_1$$

dengan:

$$c_1 = -6,13 \times 10^{-1} + 1,54 x_{2i} - 4,69 x_{3i} - 1,32 x_{4i} - 6,25 \times 10^{-4} x_{5i} - 1,76 \times 10^{-1} x_{6i} - 1,64 (x_{2i} - 2,68)_+ + 1,41 \times 10^{-1} (x_{2i} - 6,05)_+ + 1,64 (x_{2i} - 6,40)_+ + 4,05 (x_{3i} - 6,57)_+ + 1,09 \times 10^1 (x_{3i} - 10,18)_+ - 1,49 \times 10^1 (x_{3i} - 10,56)_+ + 1,20 (x_{4i} - 3,08)_+ - 3,82 (x_{4i} - 6,66)_+ + 1,03 \times 10^1 (x_{4i} - 7,05)_+ - 2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)} \right] y_i$$

Ditampilkan fungsi *truncated* untuk variabel X<sub>1</sub> sebagai berikut:

$$\hat{y} = \begin{cases} 0,56x_{1i} + c_1 & x_1 \leq 64,37 \\ 42,03 - 0,08x_{1i} + c_1 & 64,37 < x_1 \leq 71,54 \\ -247,70 + 3,96x_{1i} + c_1 & 71,54 < x_1 \leq 72,30 \\ 257,67 - 3,03x_{1i} + c_1 & x_1 > 72,30 \end{cases}$$

Interpretasi nya adalah:

- a. Kabupaten Hulu Sungai Utara merupakan kabupaten dengan angka harapan hidup yang kurang dari sama dengan 64,37 tahun.
- b. Kabupaten Hulu Sungai Tengah, Barito Kuala, Hulu Sungai Selatan, Katingan, Banjar, Barito Selatan, Balangan, Kayong Utara, Pulang Pisau, Barito Timur, Sambas, Kapuas, Kota Baru, Tanah Laut, Seruyan, Lamandau, Murung Raya, Tanah Bumbu, Kotawaringin Timur, Tapin, Kubu Raya, Tabalong, Gunung Mas, Kotawaringin Barat, Mempawah, Ketapang, Sanggau, Penajam Paser Utara, Sekadau, Nunukan, Barito Utara, Sintang, Tana Tidung, Malinau, Kota Singkawang, dan Sukamara merupakan Kabupaten/Kota dengan angka harapan hidup yang berada di rentang nilai lebih dari 64,37 tahun hingga kurang dari sama dengan 71,54 tahun.
- c. Kabupaten Mahakam Ulu, Kota Banjar Baru, Berau, Kutai Kartanegara, Kapuas Hulu dan Paser merupakan Kabupaten/Kota dengan angka harapan hidup yang berada di rentang nilai lebih dari 71,54 tahun hingga kurang dari sama dengan 72,30 tahun.
- d. Selanjutnya Kabupaten Landak, Bulungan, Melawai, Kutai Barat, Kutai Timur, Kota Palanga Raya, Bengkayang, Tarakan, Samarinda dan Balikpapan merupakan Kabupaten/Kota yang memiliki angka harapan hidup lebih dari 72,30 tahun.

**b) Model untuk variabel Tingkat Pengangguran Terbuka (X<sub>2</sub>)**

Akan diasumsikan data selain tingkat pengangguran terbuka (X<sub>2</sub>) di notasikan sebagai c<sub>2</sub>, sehingga diperoleh model umumnya:

$$\hat{y} = 1,54x_{2i} - 1,64(x_{2i} - 2,68)_+ + 1,41 \times 10^{-1}(x_{2i} - 6,05)_+ + 1,64(x_{2i} - 6,40)_+ + c_2$$

dengan:



$$\begin{aligned}
c_2 = & -6,13 \times 10^{-1} + 5,66 \times 10^{-1} x_{1i} - 4,69 x_{3i} - 1,32 x_{4i} - 6,25 \times 10^{-4} x_{5i} - 1,76 \times 10^{-1} x_{6i} - \\
& 6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ + 4,05 (x_{3i} - 6,57)_+ + \\
& 1,09 \times 10^1 (x_{3i} - 10,18)_+ - 1,49 \times 10^1 (x_{3i} - 10,56)_+ + 1,20 (x_{4i} - 3,08)_+ - 3,82 (x_{4i} - 6,66)_+ + \\
& 1,03 \times 10^1 (x_{4i} - 7,05)_+ - 2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + \\
& 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + \\
& 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K \left( \frac{t-t_i}{1,09} \right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K \left( \frac{t-t_i}{1,09} \right)} \right] y_i
\end{aligned}$$

Ditampilkan fungsi *truncated* untuk variabel  $X_2$  sebagai berikut:

$$\hat{y} = \begin{cases} 1,54x_{2i} + c_2 & x_2 \leq 2,68 \\ 4,39 - 0,10x_{2i} + c_2 & 2,68 < x_2 \leq 6,05 \\ 3,54 + 0,04x_{2i} + c_2 & 6,05 < x_2 \leq 6,40 \\ -6,95 + 1,68x_{2i} + c_2 & x_2 > 6,40 \end{cases}$$

Interpretasi nya adalah:

- Kabupaten Pulang Pisau, Gunung Mas, Balangan, Lamandau, Melawi, Kapuas Hulu, Barito Kuala, Hulu Sungai Selatan dan Bulungan merupakan Kabupaten/Kota dengan tingkat pengangguran terbuka yang kurang dari sama dengan 2,68%.
- Kabupaten Bengkayang, Landak, Hulu Sungai Tengah, Banjar, Kotawaringin Barat, Sanggau, Sekadau, Sintang, Tabalong, Murung Raya, Tanah Laut, Hulu Sungai Utara, Barito Timur, Sambas, Mahakam Ulu, Tapin, Kayong Utara, Kapuas, Barito Selatan, Barito Utara, Sukamara, Seruyan, Ketapang, Kotawaringin Timur, Paser, Katingan, Kota Baru, Kota Banjar Baru, Kutai Barat, Berau, Mempawah, Tana Tidung, Kutai Timur, Kota Tarakan, Kubu Raya, Kota Palangka Raya, Kota Samarinda, Kutai Kartanegara dan Kabupaten Malinau merupakan Kabupaten/Kota dengan tingkat pengangguran terbuka yang berada di rentang nilai lebih dari 2,68% hingga kurang dari sama dengan 6,05%.
- Kabupaten Tanah Bumbu dan Penajam Paser Utara merupakan Kabupaten dengan tingkat pengangguran terbuka yang berada di rentang nilai lebih 6,05% hingga kurang dari sama dengan 6,40%.

d. Selanjutnya Kota Singkawang, Nunukan dan Balikpapan merupakan Kabupaten/Kota yang memiliki tingkat pengangguran terbuka lebih dari 6,40%.

**c) Model untuk variabel Rata-Rata Lama Sekolah (X<sub>3</sub>)**

Akan diasumsikan data selain rata-rata lama sekolah (X<sub>3</sub>) di notasikan sebagai c<sub>3</sub>, sehingga diperoleh model umumnya:

$$\hat{y} = -4,69x_{3i} + 4,05(x_{3i} - 6,57)_+ + 1,09 \times 10^1(x_{3i} - 10,18)_+ - 1,49 \times 10^1(x_{3i} - 10,56)_+ + c_3$$

dengan:

$$c_3 = -6,13 \times 10^{-1} + 5,66 \times 10^{-1}x_{1i} + 1,54x_{2i} - 1,32x_{4i} - 6,25 \times 10^{-4}x_{5i} - 1,76 \times 10^{-1}x_{6i} - 6,53 \times 10^{-1}(x_{1i} - 64,37)_+ + 4,05(x_{1i} - 71,54)_+ - 6,99(x_{1i} - 72,30)_+ - 1,64(x_{2i} - 2,68)_+ + 1,41 \times 10^{-1}(x_{2i} - 6,05)_+ + 1,64(x_{2i} - 6,40)_+ + 1,20(x_{4i} - 3,08)_+ - 3,82(x_{4i} - 6,66)_+ + 1,03 \times 10^1(x_{4i} - 7,05)_+ - 2,49 \times 10^{-2}(x_{5i} - 34,00)_+ + 1,22 \times 10^{-1}(x_{5i} - 77,03)_+ + 3,34 \times 10^{-3}(x_{5i} - 81,56)_+ + 1,83 \times 10^{-1}(x_{6i} - 19,47)_+ - 2,58 \times 10^{-1}(x_{6i} - 77,82)_+ + 5,02 \times 10^{-1}(x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)} \right] y_i$$

Ditampilkan fungsi *truncated* untuk variabel X<sub>3</sub> sebagai berikut:

$$\hat{y} = \begin{cases} -4,69x_{3i} + c_3 & x_3 \leq 6,57 \\ -26,60 - 0,64x_{3i} + c_3 & 6,57 < x_3 \leq 10,18 \\ -137,56 + 10,26x_{3i} + c_3 & 10,18 < x_3 \leq 10,56 \\ 19,78 - 4,64x_{3i} + c_3 & x_3 > 10,56 \end{cases}$$

Interpretasi nya adalah:

- a. Kabupaten Kayong Utara dan Bengkayang merupakan Kabupaten/Kota dengan rata-rata lama sekolah yang kurang dari sama dengan 6,57 tahun.
- b. Kabupaten Sekadau, Melawi, Sambas, Mempawah, Kubu Raya, Sintang, Sanggau, Balangan, Landak, Kota Baru, Ketapang, Banjar, Barito Kuala, Hulu Sungai Utara, Tanah Laut, Murung Raya, Kapuas Hulu, Kapuas, Tapin, Tanah Bumbu, Kota Singkawang, Hulu Sungai Selatan, Hulu Sungai Tengah, Nunukan, Mahakam Ulu, Sukamara, Seruyan, Pulang Pisau, Kotawaringin Timur, Penajam Paser Utara, Kutai Barat, Lamandau,

Kotawaringin Barat, Tana Tidung, Paser, Tabalong, Barito Utara, Katingan, Barito Selatan, Bulungan, Gunung Mas, Malinau, Kutai Kartanegara, Kutai Timur, Barito Timur, Berau dan Tarakan merupakan Kabupaten/Kota dengan rata-rata lama sekolah yang berada di rentang nilai lebih dari 6,57 tahun hingga kurang dari sama dengan 10,18 tahun.

- c. Kota Samarinda merupakan Kota dengan rata-rata lama sekolah yang berada di rentang nilai lebih 10,18 tahun hingga kurang dari sama dengan 10,56 tahun.
- d. Selanjutnya Kota Balikpapan, Banjar Baru dan Palangka Raya merupakan Kabupaten/Kota yang memiliki rata-rata lama sekolah lebih dari 10,56 tahun.

**d) Model untuk variabel Laju Pertumbuhan Ekonomi (X<sub>4</sub>)**

Akan diasumsikan data selain laju pertumbuhan ekonomi (X<sub>4</sub>) di notasikan sebagai c<sub>4</sub>, sehingga diperoleh model umumnya:

$$\hat{y} = -1,32x_{4i} + 1,20(x_{4i} - 3,08)_+ - 3,82(x_{4i} - 6,66)_+ + 1,03 \times 10^1(x_{4i} - 7,05)_+ + c_4$$

dengan:

$$\begin{aligned} c_4 = & -6,13 \times 10^{-1} + 5,66 \times 10^{-1} x_{1i} + 1,54 x_{2i} - 4,69 x_{3i} - 6,25 \times 10^{-4} x_{5i} - 1,76 \times 10^{-1} x_{6i} - \\ & 6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ - 1,64 (x_{2i} - 2,68)_+ + \\ & 1,41 \times 10^{-1} (x_{2i} - 6,05)_+ + 1,64 (x_{2i} - 6,40)_+ + 4,05 (x_{3i} - 6,57)_+ + 1,09 \times 10^1 (x_{3i} - 10,18)_+ - \\ & 1,49 \times 10^1 (x_{3i} - 10,56)_+ - 2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + \\ & 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + \\ & 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K \left( \frac{t-t_i}{1,09} \right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K \left( \frac{t-t_i}{1,09} \right)} \right] y_i \end{aligned}$$

Ditampilkan fungsi *truncated* untuk variabel X<sub>4</sub> sebagai berikut:

$$\hat{y} = \begin{cases} -1,32x_{4i} + c_4 & x_4 \leq 3,08 \\ -3,69 - 0,12x_{4i} + c_4 & 3,08 < x_4 \leq 6,66 \\ 21,74 - 3,94x_{4i} + c_4 & 6,66 < x_4 \leq 7,05 \\ -50,86 - 6,36x_{4i} + c_4 & x_4 > 7,05 \end{cases}$$

Interpretasi nya adalah:

- a. Kabupaten Penajam Paser Utara dan Balangan merupakan Kabupaten/Kota dengan laju pertumbuhan ekonomi yang kurang dari sama dengan 3,08%.
- b. Kabupaten Kutai Kartanegara, Tabalong, Paser, Tanah Bumbu, Sanggau, Tanah Laut, Banjar, Bulungan, Kota Singkawang, Balikpapan, Tana Tidung, Barito Selatan, Kota Samarinda, Kayong Utara, Kota Baru, Sambas, Landak, Tapin, Seruyan, Kapuas Hulu, Bengkayang, Barito Utara, Melawi, Mahakam Ulu, Barito Timur, Sintang, Kubu Raya, Berau, Kutai Barat, Hulu Sungai Tengah, Barito Kuala, Hulu Sungai Utara, Kotawaringin Barat, Katingan, Mempawah, Sekadau, Murung Raya, Sukamara, Hulu Sungai Selatan, Malinau dan Pulang Pisau merupakan Kabupaten/Kota dengan laju pertumbuhan ekonomi yang berada di rentang nilai lebih dari 3,08% hingga kurang dari sama dengan 6,66%.
- c. Kabupaten Nunukan, Lamandau dan Kota Banjar Baru merupakan Kabupaten/Kota dengan laju pertumbuhan ekonomi yang berada di rentang nilai lebih 6,66% hingga kurang dari sama dengan 7,05%.
- d. Selanjutnya Kotawaringin Timur, Kapuas, Kota Palangka Raya, Gunung Mas, Tarakan, Kutai Timur dan Ketapang merupakan Kabupaten/Kota yang memiliki laju pertumbuhan ekonomi lebih dari 7,05%.

**e) Model untuk variabel Persentase RT dengan Akses Sanitasi Layak (X<sub>5</sub>)**

Akan diasumsikan data selain persentase RT dengan akses sanitasi layak (X<sub>5</sub>)

di notasikan sebagai  $c_5$ , sehingga diperoleh model umumnya:

$$\hat{y} = -6,25 \times 10^{-4} x_{5i} - 2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ + c_5$$

dengan:

$$c_5 = -6,13 \times 10^{-1} + 5,66 \times 10^{-1} x_{1i} + 1,54 x_{2i} - 4,69 x_{3i} - 1,32 x_{4i} - 1,76 \times 10^{-1} x_{6i} - 6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ - 1,64 (x_{2i} - 2,68)_+ + 1,41 \times 10^{-1} (x_{2i} - 6,05)_+ + 1,64 (x_{2i} - 6,40)_+ + 4,05 (x_{3i} - 6,57)_+ + 1,09 \times 10^1 (x_{3i} - 10,18)_+ - 1,49 \times 10^1 (x_{3i} - 10,56)_+ + 1,20 (x_{4i} - 3,08)_+ - 3,82 (x_{4i} - 6,66)_+ + 1,03 \times 10^1 (x_{4i} - 7,05)_+ + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)} \right] y_i$$

Ditampilkan fungsi *truncated* untuk variabel  $X_5$  sebagai berikut:

$$\hat{y} = \begin{cases} -6,25 \times 10^{-4} x_{5i} + c_5 & x_5 \leq 34,00 \\ 0,85 - 0,03x_{5i} + c_5 & 34,00 < x_5 \leq 77,03 \\ -8,55 + 0,09x_{5i} + c_5 & 77,03 < x_5 \leq 81,56 \\ -8,82 + 0,10x_{5i} + c_5 & x_5 > 81,56 \end{cases}$$

Interpretasi nya adalah:

- a. Kabupaten Murung Raya dan Kapuas merupakan Kabupaten/Kota dengan persentase rumah tangga dengan akses sanitasi layak yang kurang dari sama dengan 34%.
  - b. Kabupaten Kotawaringin Timur, Mahakam Ulu, Barito Timur, Barito Selatan, Landak, Barito Kuala, Sintang, Kapuas Hulu, Melawi, Katingan, Sanggau, Kayong Utara, Kutai Barat, Kutai Kartanegara, Gunung Mas, Sekadau, Lamandau, Tanah Laut, Penajam Paser Utara, Kota Baru, Seruyan, Paser, Nunukan, Kubu Raya, Malinau, Kotawaringin Barat, Mempawah, Tapin, Kota Tarakan, Banjar dan Ketapang merupakan Kabupaten/Kota dengan persentase rumah tangga dengan akses sanitasi layak yang berada di rentang nilai lebih dari 34% hingga kurang dari sama dengan 77,03%.
  - c. Kabupaten Bengkayang, Tanah Bumbu, Hulu Sungai Tengah, Kutai Timur dan Hulu Sungai Selatan merupakan Kabupaten/Kota dengan persentase rumah tangga dengan akses sanitasi layak yang berada di rentang nilai lebih 77,03% hingga kurang dari sama dengan 81,56%.
  - d. Selanjutnya Kabupaten Sambas, Berau, Balangan, Tana Tidung, Hulu Sungai Utara, Sukamara, Balikpapan, Banjar Baru, Bulungan, Palangka Raya, Singkawang, Tabalong dan Kota Samarinda merupakan Kabupaten/Kota yang memiliki persentase rumah tangga dengan akses sanitasi layak lebih dari 81,56%.
- f) Model untuk variabel Persentase RT dengan Akses Air Minum Layak ( $X_6$ )**  
 Akan diasumsikan data selain persentase RT dengan akses air minum layak ( $X_6$ ) di notasikan sebagai  $c_6$ , sehingga diperoleh model umumnya:

$$\hat{y} = -1,76 \times 10^{-1} x_{6i} + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + c_6$$

dengan:

$$c_6 = -6,13 \times 10^{-1} + 5,66 \times 10^{-1} x_{1i} + 1,54 x_{2i} - 4,69 x_{3i} - 1,32 x_{4i} - 6,25 \times 10^{-4} x_{5i} -$$

$$6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ - 1,64 (x_{2i} - 2,68)_+ +$$

$$1,41 \times 10^{-1} (x_{2i} - 6,05)_+ + 1,64 (x_{2i} - 6,40)_+ + 4,05 (x_{3i} - 6,57)_+ + 1,09 \times 10^1 (x_{3i} - 10,18)_+ -$$

$$1,49 \times 10^1 (x_{3i} - 10,56)_+ + 1,20 (x_{4i} - 3,08)_+ - 3,82 (x_{4i} - 6,66)_+ + 1,03 \times 10^1 (x_{4i} - 7,05)_+ -$$

$$2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ +$$

$$\frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K \left( \frac{t-t_i}{1,09} \right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K \left( \frac{t-t_i}{1,09} \right)} \right] y_i$$

Ditampilkan fungsi *truncated* untuk variabel  $X_6$  sebagai berikut:

$$\hat{y} = \begin{cases} -1,76 \times 10^{-1} x_{6i} + c_6 & x_6 \leq 19,47 \\ -3,56 + 0,01 x_{6i} + c_6 & 19,47 < x_6 \leq 77,82 \\ 16,52 - 0,25 x_{6i} + c_6 & 77,82 < x_6 \leq 83,96 \\ -25,63 + 0,25 x_{6i} + c_6 & x_6 > 83,96 \end{cases}$$

Interpretasi nya adalah:

- Kabupaten Mahakam Ulu, Kutai Kartanegara, Tanah Bumbu, Kota Banjar Baru, Kutai Timur, Penajam Paser Utara, Sintang dan Berau merupakan Kabupaten/Kota dengan persentase rumah tangga dengan akses air minum layak yang kurang dari sama dengan 19,47%.
- Kabupaten Melawi, Kota Samarinda, Tanah Laut, Kota Baru, Paser, Balikpapan, Ketapang, Kutai Barat, Tabalong, Sekadau, Banjar, Tapin, Hulu Sungai Selatan, Landak, Hulu Sungai Tengah, Barito Kuala, Kapuas Hulu, Sanggau, Murung Raya, Singkawang, Barito Timur, Barito Selatan, Balangan, Kapuas, Hulu Sungai Utara, Seruyan, Gunung Mas, Barito Utara, Malinau, Bengkayang, Lamandau, Mempawah, Kotawaringin Barat dan Katingan merupakan Kabupaten/Kota dengan persentase rumah tangga dengan akses air minum layak yang berada di rentang nilai lebih dari 19,47% hingga kurang dari sama dengan 77,82%.
- Kabupaten Bulungan, Tana Tidung, Pulang Pisau dan Kotawaringin Timur merupakan Kabupaten/Kota dengan persentase rumah tangga dengan akses

air minum layak yang berada di rentang nilai lebih 77,82% hingga kurang dari sama dengan 83,96%.

- d. Selanjutnya Kabupaten Kayong Utara, Sukamara, Kubu Raya, Sambas, Nunukan, Tarakan dan Kota Palangka Raya merupakan Kabupaten/Kota yang memiliki persentase rumah tangga dengan akses air minum layak lebih dari 83,96%.

#### 4.4 Pembahasan

Pada Sub-bab 4.1 telah dilakukan kajian mengenai model regresi nonparametrik estimator campuran *spline truncated* dan Kernel Gaussian, dengan model umum regresi nonparametrik telah dituliskan pada Persamaan (4.11). Selanjutnya didapatkan bentuk estimasi dari estimator campuran regresi nonparametrik *spline truncated* dan Kernel Gaussian sebagai berikut:

$$\begin{aligned}\hat{\mu}_{\phi,\alpha}(x,t) &= \sum_{p=1}^q \hat{f}_p(x_{pi}) + \hat{h}_{\alpha}(t) \\ &= \mathbf{A}(\phi,\alpha)\vec{Y} + \mathbf{D}(\alpha)\vec{Y} \\ &= \mathbf{B}(\phi,\alpha)\vec{Y}\end{aligned}\tag{4.34}$$

Berdasarkan Persamaan (4.34) diketahui jika,  $\mathbf{A}(\phi,\alpha)$  merupakan matriks komponen *spline truncated* dan  $\mathbf{D}(\alpha)$  merupakan matriks komponen Kernel Gaussian. Matriks  $\mathbf{B}(\phi,\alpha)$  yang merupakan penjumlahan dari matriks  $\mathbf{A}(\phi,\alpha)$  dan  $\mathbf{D}(\alpha)$  sangat tergantung pada penentuan titik knot dan *bandwidth* yang optimal. Selanjutnya  $\hat{\mu}_{\phi,\alpha}(x,t)$  merupakan hasil estimasi kurva regresi dari estimator campuran yang merupakan nilai ekspektasi atau rata-rata.

Pada Sub-bab 4.2 telah dilakukan kajian simulasi, dimana akan dibandingkan kinerja metode pemilihan titik knot dan *bandwidth* yang optimal menggunakan metode CV, GCV dan UBR pada model regresi nonparametrik estimator campuran *spline truncated* dan Kernel Gaussian. Kajian simulasi yang dilakukan ini masih terbatas, dimana *error* yang dibangkitkan hanya mengikuti distribusi Gaussian (Normal). Dicobakan beberapa situasi/kondisi yang diatur, dimana variasi ukuran sampel yang dicobakan adalah 25, 50, 100 dan 200 serta variasi dari varians *error*  $\sigma^2$  yang dicobakan adalah 0,05; 0,5 dan 1. Berdasarkan hasil kajian simulasi,

diketahui jika metode GCV lebih sesuai untuk menduga titik knot dan *bandwidth* yang optimal pada model regresi nonparametrik estimator campuran untuk setiap kombinasi variasi ukuran sampel dan varians.

Pada Sub-bab 4.3 telah dilakukan pemodelan regresi nonparametrik estimator campuran *spline truncated* dan Kernel Gaussian pada data aplikasi, yaitu persentase penduduk miskin Kabupaten/Kota di pulau Kalimantan, dimana metode pemilihan titik knot dan *bandwidth* yang digunakan adalah CV, GCV dan UBR. Berdasarkan hasil analisis, diketahui jika model yang terbaik adalah model regresi nonparametrik estimator campuran dengan tiga titik knot, dimana metode pemilihan titik knot dan *bandwidth* yang optimal menggunakan GCV.

Pada kasus data aplikasi, observasi penelitian yang digunakan hanya 53 Kabupaten/Kota seperti yang telah dijelaskan pada Batasan Masalah. Setelah dilakukan identifikasi awal pada data, diketahui jika terdapat 3 Kabupaten/Kota (Bontang, Banjarmasin dan Pontianak) yang merupakan data pencilan pada variabel prediktor, sehingga ke-3 Kabupaten/Kota tersebut tidak dijadikan unit observasi pada penelitian ini. Dampak yang terjadi ketika digunakan observasi yang lengkap (56 Kabupaten/Kota) adalah didapatkan hasil taksiran variabel respon ke-*i* yang tidak mendekati nilai aktualnya sedemikian sehingga cenderung untuk menurunkan nilai Koefisien Determinasi atau  $R^2$  untuk setiap metode pemilihan titik knot dan *bandwidth* yang digunakan pada penelitian ini yaitu CV, GCV dan UBR. Hal ini terjadi karena  $\hat{\mu}_{\phi, \alpha}(x, t)$  atau hasil estimasi kurva regresi dari estimator campuran pada Persamaan (4.34) adalah nilai ekspektasi atau rata-rata, dimana seperti yang kita ketahui jika rata-rata sangat sensitif terhadap data pencilan.



## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang dilakukan pada Bab 4, maka dapat diperoleh kesimpulan sebagai berikut:

1. Model regresi nonparametrik dengan estimator campuran *spline truncated* dan kernel Gaussian dapat dituliskan:

$$y_i = \mu(x_i, t_i) + \varepsilon_i \text{ dengan } \mu(x_i, t_i) = \sum_{p=1}^q f_p(x_{pi}) + h_\alpha(t_i)$$

Sehingga apabila dilakukan substitusi akan didapatkan

$$y_i = \left[ \sum_{p=1}^q f_p(x_{pi}) + h_\alpha(t_i) \right] + \varepsilon_i$$

Kemudian dapat di sajikan dalam bentuk matriks menjadi:

$$\vec{Y} = \mathbf{X}(\phi)\vec{\beta} + \mathbf{D}(\alpha)\vec{Y} + \vec{\varepsilon}$$

Estimasi parameter dari  $\vec{\beta}$  dapat diperoleh dengan menggunakan metode

*Ordinary Least Squares* (OLS). Maka akan di dapatkan estimator untuk  $\hat{\vec{\beta}}$

sebagai berikut:  $\hat{\vec{\beta}} = [\mathbf{X}(\phi)^T \mathbf{X}(\phi)]^{-1} \mathbf{X}(\phi)^T [\mathbf{I} - \mathbf{D}(\alpha)]\vec{Y}$ .

2. Berdasarkan data simulasi dengan *error* yang mengikuti distribusi Normal serta dilakukan kombinasi variasi ukuran sampel dan varians *error*, didapatkan hasil bahwa secara keseluruhan rata-rata dari hasil estimasi titik knot dan *bandwidth* yang dihasilkan dari metode CV, GCV dan UBR dapat dikatakan cukup baik. Namun metode GCV memberikan performa hasil serta ketepatan yang lebih baik untuk setiap kombinasi jumlah sampel dan variasi varians yang dicobakan. Metode GCV menghasilkan titik knot dan *bandwidth* yang optimal sehingga didapatkan nilai koefisien determinasi ( $R^2$ ) yang terbesar untuk setiap kombinasi. Sehingga dapat disimpulkan jika metode GCV lebih sesuai untuk menduga titik knot dan *bandwidth* pada model regresi nonparametrik estimator campuran *spline truncated* dan kernel Gaussian.

3. Hasil perbandingan metode pemilihan titik knot dan *bandwidth* yang optimal menggunakan CV, GCV dan UBR pada data aplikasi, yaitu persentase penduduk miskin Kabupaten/Kota di Pulau Kalimantan Tahun 2019 adalah:
- Model terbaik untuk metode CV adalah model regresi nonparametrik estimator campuran dengan tiga titik knot yang menghasilkan nilai Koefisien Determinasi ( $R^2$ ) sebesar 83,80%.
  - Model terbaik untuk metode GCV adalah model regresi nonparametrik estimator campuran dengan tiga titik knot yang menghasilkan nilai Koefisien Determinasi ( $R^2$ ) sebesar 90,52%.
  - Model terbaik untuk metode UBR adalah model regresi nonparametrik estimator campuran dengan tiga titik knot yang menghasilkan nilai Koefisien Determinasi ( $R^2$ ) sebesar 79,24%.

Secara keseluruhan, dapat disimpulkan jika dengan menggunakan metode GCV dalam proses pemilihan titik knot dan *bandwidth* yang optimal pada model regresi nonparametrik estimator campuran, diperoleh hasil yang lebih baik dibandingkan dengan metode CV maupun UBR.

Adapun model regresi nonparametrik estimator campuran fungsi *spline truncated* linier dan kernel Gaussian dengan tiga titik knot menggunakan metode GCV sebagai berikut:

$$\hat{y} = -6,13 \times 10^{-1} + 5,66 \times 10^{-1} x_{1i} + 1,54 x_{2i} - 4,69 x_{3i} - 1,32 x_{4i} - 6,25 \times 10^{-4} x_{5i} - 1,76 \times 10^{-1} x_{6i} - 6,53 \times 10^{-1} (x_{1i} - 64,37)_+ + 4,05 (x_{1i} - 71,54)_+ - 6,99 (x_{1i} - 72,30)_+ - 1,64 (x_{2i} - 2,68)_+ + 1,41 \times 10^{-1} (x_{2i} - 6,05)_+ + 1,64 (x_{2i} - 6,40)_+ + 4,05 (x_{3i} - 6,57)_+ + 1,09 \times 10^1 (x_{3i} - 10,18)_+ - 1,49 \times 10^1 (x_{3i} - 10,56)_+ + 1,20 (x_{4i} - 3,08)_+ - 3,82 (x_{4i} - 6,66)_+ + 1,03 \times 10^1 (x_{4i} - 7,05)_+ - 2,49 \times 10^{-2} (x_{5i} - 34,00)_+ + 1,22 \times 10^{-1} (x_{5i} - 77,03)_+ + 3,34 \times 10^{-3} (x_{5i} - 81,56)_+ + 1,83 \times 10^{-1} (x_{6i} - 19,47)_+ - 2,58 \times 10^{-1} (x_{6i} - 77,82)_+ + 5,02 \times 10^{-1} (x_{6i} - 83,96)_+ + \frac{1}{53} \sum_{i=1}^{53} \left[ \frac{\frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)}{\frac{1}{53} \sum_{i=1}^{53} \frac{1}{1,09} K\left(\frac{t-t_i}{1,09}\right)} \right] y_i$$

Model ini memberikan nilai Koefisien Determinasi ( $R^2$ ) sebesar 90,52%. Nilai  $R^2$  ini menunjukkan bahwa variabel prediktor yang digunakan dapat menjelaskan model sebesar 90,52%.

## 5.2 Saran

Berdasarkan penelitian yang telah dilakukan, maka permasalahan yang dikaji dapat dikembangkan untuk dijadikan penelitian selanjutnya, antara lain:

1. Metode yang digunakan dalam penelitian ini mengasumsikan setiap variabel prediktor yang didekati dengan kurva *spline truncated* mempunyai jumlah knot yang sama, yaitu 1, 2, hingga 3 titik knot. Maka pada penelitian selanjutnya dapat dikembangkan dengan menggunakan sistem kombinasi titik knot, dimana untuk setiap variabel memiliki titik knot yang berbeda-beda jumlahnya.
2. Penelitian ini terbatas hanya untuk penentuan nilai estimasi titik. Pada penelitian selanjutnya dapat menggunakan pendugaan interval kepercayaan dalam proses penentuan estimasi.
3. Pada bagian kajian simulasi, *error* yang dibangkitkan terbatas hanya mengikuti distribusi Normal  $(0, \sigma^2)$ . Pada penelitian selanjutnya dapat dikembangkan jika *error* yang dibangkitkan mengikuti distribusi lain, sebagai contoh distribusi Eksponensial.
4. Perlu dilakukan kajian lebih lanjut dengan memperhatikan lokasi atau efek spasial untuk setiap Kabupaten/Kota. Pada penelitian selanjutnya dapat dikembangkan model *Nonparametric Geographically Weighted Regression*.

*(Halaman ini sengaja dikosongkan)*

## DAFTAR PUSTAKA

- Adhi, E. T., (2009), "Pelayanan Sanitasi Buruk Akar dari Kemiskinan", *Jurnal Analisis Sosial*, Vol. 14, No. 02, hal: 76-87.
- Aljuhani, K. H. & Alturk, L. I., (2014), "Modification of the Adaptive Nadaraya-Watson Kernel Regression Estimator", *Academic Journal*, Vol. 09, No .22, hal: 966-971.
- Andrews, D. W. K, (1991), "Asymptotic Optimality of Generalized  $C_L$ , Cross-Validation, and Generalized Cross-Validation in Regression with Heteroskedastic Errors", *Journal of Econometrics*, Vol. 47, hal: 359-377.
- Astuti, Dewa A. D., Srinadi, I. G. A. M. & Susilawati, M., (2018), "Pendekatan Regresi Nonparametrik dengan Menggunakan Estimator Kernel pada Data Kurs Rupiah terhadap Dollar Amerika Serikat", *E-Jurnal Matematika*, Vol. 07, No. 04, hal: 305-310.
- Aydin, D., Guneri, O. I. & Fit, A., (2016), "Choice of Bandwidth for Nonparametric Regression Models using Kernel Smoothing: A Simulation Study", *International Journal of Sciences Basic and Applied Research (IJSBAR)*, Vol. 26, No. 01, hal: 47-61.
- Bilodeau, M., (1992), "Fourier Smoother and Additive Models", *Canadian Journal of Statistics*, Vol. 20, No. 03, hal: 257-269.
- Badan Pusat Statistik, (2017), Laporan Akhir: Analisis Kriteria dan Indikator Kemiskinan Multidimensi untuk Diagnostik Kemajuan Daerah di Daerah Istimewa Yogyakarta.
- Badan Pusat Statistik (BPS), (2020), *Provinsi Kalimantan Barat dalam Angka 2020*, Kalbar: BPS.
- Badan Pusat Statistik (BPS), (2020), *Provinsi Kalimantan Timur dalam Angka 2020*, Kaltim: BPS.
- Badan Pusat Statistik (BPS), (2020), *Provinsi Kalimantan Utara dalam Angka 2020*, Kaltara: BPS.
- Badan Pusat Statistik (BPS), (2020), *Provinsi Kalimantan Selatan dalam Angka 2020*, Kalsel: BPS.

- Badan Pusat Statistik (BPS), (2020), *Provinsi Kalimantan Tengah dalam Angka 2020*, Kalteng: BPS.
- Budiantara, I. N., (2000), “Metode U, GML, CV dan GCV dalam Regresi Nonparametrik Spline”, *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 06, hal: 285-290.
- Budiantara, I. N., (2004), “Spline: Historis, Motivasi, dan Perannya dalam Regresi Nonparametrik”, *Makalah Pembicara Utama pada Konferensi Nasional Matematika XII FMIPA Universitas Udayana Denpasar, 23-27 Juli 2004*.
- Budiantara, I. N., (2009), *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*, Pidato Pengukuhan Untuk Jabatan Guru Besar Dalam Bidang Ilmu: Matematika Statistika dan Probabilitas, Pada Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Insititut Teknologi Sepuluh November, Surabaya.
- Budiantara, I. N., (2011), *Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter*, Prosiding Seminar Nasional FMIPA Undiksha, hal: 09-28.
- Budiantara, I. N., Ratna, M., Zain, I. & Wibowo, W., (2012), “Modelling the Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach”, *International Journal of Basic & Applied Sciences IJBAS-IJENS*, Vol. 12, No. 06, hal: 119-124.
- Budiantara, I. N., Ratnasari, V., Madu, R., & Zain, I., (2015), “The Combination of Spline and Kernel Estimator for Nonparametric Regression and Its Properties”, *Applied Mathematical Sciences*, Vol. 09, No. 22, hal: 6083-6094.
- Budiantara, I. N., (2019), *Regresi Nonparametrik Spline Truncated*, Surabaya: ITS Press.
- Budiantara, I. N., Ratnasari, V., Ratna, M., Wibowo, W., Afifah, N., Rahmawati, D. P., & Octavanny, M. A. D., (2019), “Modeling Percentage of Poor People in Indonesia using Kernel and Fourier Series Mixed Estimator in Nonparametric Regression”, *Revista Investigacion Operacional*, Vol. 40, No. 40, hal: 538-550.

- Chong, J., Abeysuriya, K., Hidayat, L., Sulistio, H. & Willetts, J., (2016), “Strengthening Local Governance Arrangements for Sanitation in Small Cities in Indonesia”, *Aquatic Procedia*, Vol. 06, hal: 64-73.
- Craven, P. & Wahba, G., (1979), “Smoothing Noisy Data with Spline Functions”. *Numerische Mathematics*, Vol. 31, hal: 377-403.
- Cutler, D. M. & Katz, L. F., (1991), “Macroeconomic Performance and The Disadvantaged”, *Brooking Papers on Economic Activity*, hal: 01-74.
- Damaliana, A. T., Budiantara, I. N. & Ratnasari, V., (2019), “Comparing Between mGCV and aGCV Methods to Choose the Optimal Knot Points in Semiparametric Regression with Spline Truncated Using Longitudinal Data”, *9th Annual Basic Science International Conference 2019 (BaSIC 2019)*, IOP Publishing DOI: 10.1088/1757-899X/546/3/032003.
- Devi, A. R., Budiantara, I. N. & Ratnasari, V., (2018), “Unbiased Risk and Cross-Validation Method for Selecting Optimal Knots in Multivariable Nonparametric Regression Spline Truncated”, *The 8<sup>th</sup> Annual Basic Science International Conferences*. AIP Conf. Proc. 2021, 060003-1 - 060003-9.
- Didu, S. dan Fauzi, F., (2016), “Pengaruh Jumlah Penduduk, Pendidikan dan Pertumbuhan Ekonomi terhadap Kemiskinan di Kabupaten Lebak”, *Jurnal Ilmu Ekonomi*, Vol. 06, No. 01, hal: 100-114.
- Dores, E. & Jolianis, (2014), “Pengaruh Angka Melek Huruf dan Angka Harapan Hidup terhadap Jumlah Penduduk Miskin di Provinsi Sumatera Barat”, *Journal of Economic and Economic Education*, Vol. 02, No. 02, pp. 126-133.
- Draper, N. R. & Smith, H., (1992), *Applied Regression Analysis: Third Edition*, USA: John Wiley & Sons, Inc.
- Du, P., Parmeter, C. F & Racine. J. S., (2013), “Nonparametric Kernel Regression with Multiple Predictors and Multiple Shape Constraints”, *Statistica Sinica*, Vol. 23, hal: 1347-1371.
- Eubank, R. L., (1999), *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker.
- Fatmawati, Budiantara, I. N. & Lestari, B., (2019), “Comparison of Smoothing and Truncated Spline Estimators in Estimating Blood Pressure Models”,

- International Journal of Innovation, Creativity and Change*, Vol. 05, hal: 1177-1199.
- Ferezagia, D. V., (2018), “Analisis Tingkat Kemiskinan di Indonesia”, *Jurnal Sosial Humaniora*, Vol. 01, No. 01, hal: 01-06.
- Fox, J., (2008), *Applied Regression Analysis and Generalized Linier Models*, London: Sage Publications.
- Ginting, A. M. & Dewi, G. P., (2013), “Pengaruh Pertumbuhan Ekonomi dan Pertumbuhan Sektor Keuangan terhadap Pengurangan Kemiskinan di Indonesia”, *Jurnal Ekonomi & Kebijakan Publik*, Vol. 04, No. 02, hal: 117-130.
- Green, P. J. & Silverman, B. W., (1994), *Nonparametric Regression and Generalized Linear Models*, London: Chapman & Hall.
- Gu, C., Heckman, N. & Wahba, G., (1992), “A Note on Generalized Cross-Validation with Replicates”, *Statistics & Probability Letters*, Vol. 14, hal: 283-287.
- Hadi, A., (2017), “Pengaruh Rata-Rata Lama Sekolah Kabupaten/Kota terhadap Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Timur Tahun 2017”, *Media Trend: Kajian Ekonomi dan Studi Pembangunan*, Vol. 14, No. 02, pp. 148-153.
- Hardle, W., (1990), *Applied Nonparametric Regression*, New York: Cambridge University Press.
- Hardle, W., (1994), *Applied Nonparametric Regression*, Berlin: Humboldt-Universitat zu Berlin.
- Hartanto, T. B. & Masjkuri, S. U., (2017), “Analisis Pengaruh Jumlah Penduduk, Pendidikan, Upah Minimum dan Produk Domestik Regional Bruto (PDRB) terhadap Jumlah Pengangguran di Kabupaten dan Kota, Provinsi Jawa Timur Tahun 2010-2014”, *Jurnal Ilmu Ekonomi Terapan*, Vol. 02, No. 1, hal: 01-11.
- Hastie, T. J. & Tibshirani, R. J., (1990), *Generalized Additive Models*, New York: Chapman and Hall.
- Hidayat, R., Budiantara, I. N., Otok, B. W. & Ratnasari, V., (2019), “Kernel-Spline Estimation of Additive Nonparametric Regression Model”, *9<sup>th</sup> Annual Basic*



- Science International Conference 2019, IOP Conf. Series: Materials Science and Engineering*, hal: 01-09.
- Jacobus, E. H., Kindangen, P. & Walewangko, E. N., (2018), “Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan Rumah Tangga di Sulawesi Utara”, *Jurnal Pembangunan Ekonomi dan Keuangan Daerah*, Vol. 19, No. 07, hal: 01-16.
- Kartasasmita, G., (1996), *Pembangunan untuk Rakyat: Memadukan Pertumbuhan dan Pemerataan*, Jakarta: CIDES.
- Kayri, M. & Zirhlioglu, G., (2009), “Kernel Smoothing Function and Choosing Bandwidth for Nonparametric Regression Methods”, *Ozean Journal of Applied Sciences*, Vol. 2, hal: 49-54.
- Klemela, J., (2014), *Multivariate Nonparametric Regression and Visualization: With R and Applications to Finance*, New Jersey: John Wiley & Sons, Inc.
- Lin, Y. & Zhang, H. H., (2006), “Component Selection and Smoothing in Multivariate Nonparametric Regression”, *The Annals of Statistics*, Vol. 34, No. 05, hal: 2272-2297.
- Maizunati, N. A., (2017), “Pengaruh Dimensi Sosial dan Lingkungan terhadap Tingkat Konsumsi Pangan pada Penduduk Miskin Perkotaan di Jawa Tengah”, *Jurnal Riset Ekonomi Pembangunan*, Vol. 02, No. 3, pp. 227-248.
- Merdekawati, I. P., & Budiantara, I. N., (2013), “Pemodelan Regresi Spline Truncated Multivariabel pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah”, *Jurnal Sains dan Seni ITS*, Vol. 02, No. 01, hal: 19-24.
- Mils. A. & Gilson, L., (1990). *Ekonomi Kesehatan untuk Negara Sedang Berkembang*, Jakarta: Dian Rakyat.
- Montoya, E. L., Ulloa, N., & Miler, V., (2014), “A Simulation Study Comparing Knot Selection Methods with Equality Spaced Knots in a Penalized Regression Spline”, *International Journal of Statistics and Probability*, Vol. 03, No. 03, hal: 96-110.
- Morton, J., & Silverberg, L., (1999), “Fourier Series of Half-range Functions by Smooth Extension”, *Applied Mathematical Modelling*, Vol. 33, No. 02, hal: 812-821.

- Mubarak, R. & Budiantara, I. N., (2012), “Analisis Regresi *Spline* Multivariabel untuk Pemodelan Kematian Penderita Demam Berdarah Dengue (DBD) di Jawa Timur”, *Jurnal Sains dan Seni ITS*, Vol. 01, No. 01, hal: 224-229.
- Muda, R., Koleangan, R. & Kalangi, J. B., (2019), “Pengaruh Angka Harapan Hidup, Tingkat Pendidikan dan Pengeluaran Perkapita terhadap Pertumbuhan Ekonomi di Sulawesi Utara Tahun 2003-2017”, *Jurnal Berkala Ilmiah Efisiensi*, Vol. 19, No. 01, hal: 44-55.
- Mungkasa, O., (2004), *Sekilas Kondisi Air Minum dan Sanitasi di Indonesia*, Jakarta: Media Percik.
- Nadaraya, E. A., (1989), *Nonparametric Estimation of Probability Densities and Regression Curves*, London: Kluwer Academic Publisher.
- Ningrum, S. S., (2017), “Analisis Pengaruh Tingkat Pengangguran Terbuka, Indeks Pembangunan Manusia dan Upah Minimum terhadap Jumlah Penduduk Miskin di Indonesia Tahun 2011-2015”, *Jurnal Ekonomi dan Pembangunan*, Vol. 15, No. 02, pp. 184-192.
- Nurjanah, F., Utami, T. W. & Nur, I. M., (2015), “Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Pola Data Curah Hujan di Kota Semarang”, *Statistika*, Vol. 03, No. 02, hal: 08-14.
- Nyompa, S., Maru, R. & Amal, (2016), “Analisis Kepadatan Penduduk dengan Tingkat Kemiskinan di Kota Makassar”, *Prosiding Seminar Nasional LP2M UNM*, hal: 902-906.
- Prahotama, A., Suparti & Utami, T. W., (2018), “Modelling Fourier Regression for Time Series Data with Case Study Modelling Inflation in Foods Sector in Indonesia”, *International Conference on Mathematics: Pure, Applied and Computation*, IOP Conf. Series: Journal of Physics, hal: 01-09.
- Pratiwi, L. P. S. P., (2017), “Perbandingan Metode Cross Validation dan Generalized Cross Validation dalam Regresi Nonparametrik Birespon *Spline*”. *Jurnal Varian*, Vol. 01, No. 01, hal: 43-53.
- Purnomo, A. A. S. I., (2016), *Estimator Campuran Kernel dan Spline Truncated Linier Multivariabel dalam Regresi Nonparametrik (Studi Kasus: Model Rata-Rata Lama Sekolah di Provinsi Jawa Tengah)*, Tesis, Surabaya: Institut Teknologi Sepuluh November.

- Putra, H. S. P. & Rianto, N., (2016), “Pengaruh Akses Air Bersih terhadap Kemiskinan di Indonesia: Pengujian Data Rumah Tangga”, *Jurnal Sosial Ekonomi Pekerjaan Umum*, hal: 65-76.
- Rahim, F., Budiantara, I. N., & Permatasari, E. A., (2019), “Spline Truncated Nonparametric Regression Modelling for Maternal Mortality Rate in East Java”, *INFERENSI*, Vol. 02, hal: 39-44.
- Ratnasari, V., Budiantara, I. N., Zain, I. Ratna, M. & Mariati, N. P. A. M., (2015), “Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Model: Data of Poverty in Papua, Indonesia”, *International Journal of Basic & Applied Sciences IJBAS-IJENS*, Vol. 15, No. 04, hal: 09-14.
- Ratnasari, V., Budiantara, I. N., Ratna, M., & Zain, I., (2016), “Estimation of Nonparametric Regression Curve Using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel”, *Global Journal of Pure and Applied Mathematics*, Vol. 12, No. 06, hal: 5047:5057.
- Rizki, B. & Saleh, S., (2007), “Keterkaitan Akses Sanitasi dan Tingkat Kemiskinan di Provinsi Jawa Tengah”, *Jurnal Ekonomi Pembangunan*, Vol. 12, No. 03, hal: 223-233.
- Safuridar & Damayanti, M., (2018), “Analisis Pengaruh Pendidikan dan PBRB per Kapita terhadap Jumlah Penduduk Miskin di Provinsi Aceh”, *Jurnal Samudra Ekonomi dan Bisnis*, Vol. 09, No. 02, pp. 180-187.
- Saputro, D. R. S., Demu, K. R. & Widyaningsih, P., (2018), “Nonparametric Truncated Spline Regression Model on the Data of Human Development Index (HDI) in Indonesia”, *2<sup>nd</sup> International Conference on Statistics, Mathematics, Teaching and Research*, IOP Conf. Series: Journal of Physics, hal: 01-04.
- Sari, S. U. R., (2016), *Comparison of Multivariable Spline Nonparametric Regression Model using Generalized Cross Validation (GCV) and Unbiased Risk (UBR) in Selecting the Optimal Knot Points (Case Study: Data of Maternal Mortality Rate in East Java)*, Tesis, Surabaya: Institut Teknologi Sepuluh Nopember.

- Sartika, C., Balaka, M. Y. & Rumbia, W. A., (2016), “Studi Faktor-Faktor Penyebab Kemiskinan Masyarakat Desa Lohia Kabupaten Muna”, *Jurnal Ekonomi*, Vol. 01, No. 01, hal: 106-118.
- Setiawan, R. N. S., Budiantara, I. N., & Ratnasari, V., (2017), Application of Confidence Intervals for Parameters of Nonparametric Spline Truncated Regression on Index Development Gender in East Java, *IPTEK Journal of Sciences*, Vol. 02, No. 03, hal: 49-55.
- Sriningsih, M., Hatidja, D., & Prang, J. D., 2018, “Penanganan Multikolinearitas dengan Menggunakan Analisis Regresi Komponen Utama pada Kasus Impor Beras di Provinsi SULUT”, *Jurnal Ilmiah Sains*, Vol. 18, No. 01, hal: 18-24.
- Sukarsa, I. K. G. & Srinadi, I. G. A. M., 2012, “Estimator Kernel dalam Model Regresi Nonparametrik”, *Jurnal Matematika*, Vol. 02, No. 01, hal: 19-30.
- Suparti & Prahutama, A., (2016), “Pemodelan Regresi Nonparametrik Menggunakan Pendekatan Polinomial Lokal pada Beban Listrik di Kota Semarang”, *Media Statistika: Undip*, Vol. 09, No. 02, hal: 85-93.
- Suryawati, C., (2005), “Memahami Kemiskinan secara Multidimensional”, *JMPK*, Vol. 08, No. 03, hal:121-129.
- Syahputra, R., (2017), “Analisis Faktor-Faktor yang Mempengaruhi Pertumbuhan Ekonomi di Indonesia”, *Jurnal Samudra Ekonomika*, Vol. 01, No. 02, hal: 183-191.
- Tamana, N., (2018), “Akses Masyarakat Miskin terhadap Air Minum Bersih di Provinsi Jawa Timur”, *Jurnal Ilmiah: Fakultas Ekonomi dan Bisnis Universitas Brawijaya*.
- Tripena, A., (2011), “Penentuan Model Regresi Spline Terbaik”, *Prosiding Seminar Nasional Statistika Universitas Diponegoro*, hal: 92-102.
- Ustama, D. D., (2009), “Peranan Pendidikan dalam Pengentasan Kemiskinan”, *Jurnal Ilmu Administrasi dan Kebijakan Publik*, Vol. 06, No. 01, hal: 01-12.
- Wahba, G., (1990), *Spline Models for Observational Data*, Pennsylvania: SIAM.

- Wang, Y., (1997), “GRKPACK: Fitting Smoothing Spline Analysis of Variances Models to Data from Exponential Families”, *Communication in Statistics: Simulation and Computation*, Vol. 26, hal: 756-782.
- Wu, H. & Zhang, J. T., (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed Effects Modelling Approaches*, New York: John Wiley & Sons.
- Yacoub, Y., (2012), “Pengaruh Tingkat Pengangguran terhadap Tingkat Kemiskinan Kabupaten/Kota di Provinsi Kalimantan Barat”, *Jurnal EKSOS*, Vol. 08, No. 03, hal: 176-185.

*(Halaman ini sengaja dikosongkan)*

# LAMPIRAN

## Lampiran 1 Data Penelitian

Kabupaten/Kota	$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$t_1$
Sambas	8,19	68,5	3,58	6,7	5,1	81,83	91,63	84
Bengkayang	6,96	73,28	2,71	6,53	5,37	77,76	68,21	47
Landak	11,47	72,34	2,75	7,1	5,12	46,88	37,23	38
Mempawah	5,32	70,54	5,19	6,82	5,87	74,71	76,95	207
Sanggau	4,57	71,05	3,07	6,95	4,47	62	47,74	37
Ketapang	10,54	70,69	4,53	7,26	7,99	76,68	30,37	16
Sintang	9,65	71,29	3,2	6,89	5,47	57,73	16,17	19
Kapuas Hulu	9,62	72,12	2,46	7,47	5,23	57,88	44,71	9
Sekadau	6,11	71,24	3,14	6,6	5,88	64,75	33,96	37
Melawi	12,38	72,56	2,45	6,67	5,44	61,7	21,69	20
Kayong Utara	9,98	67,71	4	6	5,02	62,9	84,17	25
Kubu Raya	4,74	70,04	5,62	6,82	5,49	72,98	89,06	83
Kota Pontianak	4,88	72,41	9,13	10,14	5,03	94,05	58,03	5999
Kotawaringin Barat	4,11	70,43	3,01	8,41	5,79	73,46	77,47	29
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
Barito Utara	4,95	71,28	4,34	8,6	5,42	50,74	65,51	16
Sukamara	3,16	71,45	4,39	7,91	6,05	85,59	84,2	17
Lamandau	3,01	69,31	2,42	8,38	6,88	64,94	72,73	13
Seruyan	7,19	69,24	4,4	7,93	5,19	69,63	65,02	13
Katingan	5,02	65,62	4,78	8,66	5,81	61,94	77,72	10
Pulang Pisau	4,24	67,92	2,15	8,08	6,37	50,61	79,16	14
Gunung Mas	4,91	70,24	2,19	9,03	7,29	64,72	65,35	11
Barito Timur	6,32	68,06	3,5	9,2	5,45	42,78	55,61	33
Murung Raya	6	69,43	3,37	7,46	5,94	27,21	49,57	5
Kota Palangka Raya	3,35	73,16	5,81	11,51	7,17	89,77	99,32	122
Paser	8,95	72,28	4,55	8,54	3,88	72,48	28,09	26,31
Kutai Barat	9,09	72,57	5,08	8,34	5,58	63,53	32,28	10,83
Kutai Kartanegara	7,2	71,93	5,98	9,1	3,77	63,8	13,07	30,91
Kutai Timur	9,48	72,76	5,53	9,18	7,97	80,15	15,73	12,6
Berau	5,04	71,68	5,08	9,25	5,55	81,97	19,27	10,96
Penajam Paser Utara	7,18	71,05	6,26	8,16	2,52	67,4	15,99	55,56
Mahakam Ulu	11,25	71,56	3,69	7,89	5,44	41,61	10,26	1,36
Kota Balikpapan	2,42	74,18	7,29	10,67	4,78	86	28,55	1297,74
Kota Samarinda	4,59	73,93	5,87	10,47	4,97	92,89	23,1	1238,9
Kota Bontang	4,22	73,94	9,19	10,73	-2,18	94,73	19,94	1111,37



**Keterangan:**

Variabel	Deskripsi	Satuan
$y$	Persentase Penduduk Miskin	Persen
$x_1$	Angka Harapan Hidup	Tahun
$x_2$	Tingkat Pengangguran Terbuka	Persen
$x_3$	Rata-Rata Lama Sekolah	Tahun
$x_4$	Laju Pertumbuhan Ekonomi	Persen
$x_5$	Persentase RT dengan Akses Sanitasi Layak	Persen
$x_6$	Persentase RT dengan Akses Air Minum Layak	Persen
$t_1$	Kepadatan Penduduk	Jiwa/Km <sup>2</sup>

**Catatan:**

Kota Bontang, Kota Banjarmasin dan Kota Pontianak tidak dijadikan unit observasi dalam penelitian.

**Lampiran 2** *Syntax R Studi Simulasi Model Regresi Nonparametrik Estimator Campuran Spline Truncated dan Kernel Gaussian dengan Metode Cross-Validation (CV)*

```

#Cross Validation
library(pracma)
# Generate some data to test the implementation
n <- 50
eps <- rnorm(n, mean=0, sd = 1)
m <- function(x) x^2 * sin(x) #Kernel

f<-function(x) ((sin(pi*x^2))^4)/sin(x^2) #Spline

X1 <- runif(n,0,1)
X2 <- runif(n,0,1)
Y <- f(X1)+m(X2) + eps
xGrid <- seq(-10, 10, l = 500)

par(mfrow=c(1,2))
plot(X1,Y,main="SPLINE")
plot(X2,Y,main="KERNEL")

#matrix
m1.nn=matrix(1, nrow=n, ncol=n) #matriks 1 nxn
m1.n1=matrix(1, nrow=n) #matriks 1 nx1
mi.nn=diag(1,n,n) #matriks identitas nxn

xk=as.matrix(X2) #variabel kernel
xs=as.matrix(X1) #variabel spline kn=1 #jumlah titik knot
kn=1 #1 titik knot
ps=ncol(xs) #jumlah variabel spline
pk=ncol(as.matrix(xk)) #jumlah variabel kernel
int.sp=20 #jumlah pembagi titik knot yang diinginkan
alpha=0.05
int.kr=20 #jumlah pembagi titik bandwidth yang diinginkan

#penentuan titik knot
knot=matrix(0,int.sp,ps)
for (i in 1:ps)
{
  knot[,i]=seq(min(xs[,i]),max(xs[,i]),length.out=int.sp)
}
knot=as.matrix(knot[2:(int.sp-1),])
nknot=nrow(knot)

if (kn==1){
  knot=as.matrix(knot)
}else if (kn==2)
{
  #knot2
  nkomb=(nknot*(nknot-1)/2)
  knot2=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-1))
  {
    for (j in (i+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l])

```

```

        kk=cbind(kk,a)
    }
    knot2[v,]=kk[1,2:ncol(kk)]
    v=v+1
}
}
knot=as.matrix(knot2)
nknot=nrow(knot)
}else
{
    #knot3
    nkomb=(nknot*(nknot-1)*(nknot-2))/6)
    knot3=matrix(0,nkomb,kn*ps)
    v=1
    for (i in 1:(nknot-2))
    {
        for (j in (i+1):(nknot-1))
        {
            for (k in (j+1):nknot)
            {
                kk=0
                for (l in 1:ps)
                {
                    a=cbind(knot[i,l],knot[j,l],knot[k,l])
                    kk=cbind(kk,a)
                }
                knot3[v,]=kk[1,2:ncol(kk)]
                v=v+1
            }
        }
    }
    knot=as.matrix(knot3)
    nknot=nrow(knot)
}

#penentuan bandwidth
bw=matrix(0,int.kr,pk)
for (i in 1:pk)
{
    bw[,i]=seq(0,(max(xk[,i])-min(xk[,i])),length.out=int.kr)
}
bw=as.matrix(bw[2:(int.kr-1),])
nband=nrow(bw)

#desain matriks X(k) pada spline
MSE=matrix(0,nband*nknot)
CV=matrix(0,nband*nknot)
SSE=matrix(0,nband*nknot)
SST=matrix(0,nband*nknot)
R2=matrix(0,nband*nknot)
Normalitas=matrix(0,nband*nknot)
code=matrix(0,nband*nknot,kn*ps+pk)
o=1
for (i in 1:nknot)
{
    for (j in 1:nband)
    {
        #matrik spline
        Z=cbind(1,xs)
        a=1
        for (k in 1:ps)
        {
            for (l in 1:kn)
            {

```

```

        Z=cbind(Z, (pmax(0, xs[,k]-knot[i,a])))
        a=a+1
    }
}
sum.v.phi=0
for (k in 1:pk)
{
    v.diag=diag(xk[,k])
    V=m1.nn %*%v.diag
    z=(t(V)-V)/bw[j,k]
    K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
    K.Z=(1/bw[j,k])*K
    W.penyebut=diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
    V.phi=1/n*K.Z/W.penyebut

    #penimbang V(phi).1
    sum.v.phi=sum.v.phi+V.phi #nilai kernel untuk setiap variabel
}
# penimbang kernel gabungan
V.phi=sum.v.phi/pk #nilai kernel rata-rata

#estimasi parameter
beta=0
C=pinv(t(Z)%*%Z)%*%t(Z)%*%(m1.nn-V.phi)
beta=C%*%Y
A=Z%*%C
B=A+V.phi
yhat=B%*%Y
error=Y-yhat
residual=error
MSE[o]=n^-1*(t(error)%*%error)

SSE=sum((Y-yhat)^2)
SST=sum((Y-mean(Y))^2)
R2[o]=(1-(SSE/SST))

#Normalitas
Normalitas[o]=ks.test(residual,"pnorm",mean(residual),sqrt(var(residual)))$p.value

#Cross Validation
residual=error
A3=(rep(1,nrow(B))-diag(B))
CV[o]=sum((residual/A3)^2)/(n)
code[o,]=c(knot[i,],bw[j,])
o=o+1
}
}
optimum=cbind(code,MSE,CV,R2,Normalitas)

CVmin=optimum[order(optimum[, (kn*ps+pk+2)]),]
print(CVmin[1:1,])

```

**Lampiran 3** *Syntax* R Studi Simulasi Model Regresi Nonparametrik Estimator  
 Campuran *Spline Truncated* dan Kernel Gaussian dengan Metode  
*Generalized Cross-Validation (GCV)*

```

#Generalized Cross Validation
library(pracma)
# Generate some data to test the implementation
n <- 50
eps <- rnorm(n, mean=0, sd = 1)
m <- function(x) x^2 * sin(x) #Kernel

f<-function(x) ((sin(pi*x^2))^4)/sin(x^2) #Spline

X1 <- runif(n,0,1)
X2 <- runif(n,0,1)
Y <- f(X1)+m(X2) + eps
xGrid <- seq(-10, 10, l = 500)

par(mfrow=c(1,2))
plot(X1,Y,main="SPLINE")
plot(X2,Y,main="KERNEL")

#matrix
ml.nn=matrix(1, nrow=n, ncol=n) #matriks 1 nxn
ml.n1=matrix(1, nrow=n) #matriks 1 nx1
mi.nn=diag(1,n,n) #matriks identitas nxn

xk=as.matrix(X2) #variabel kernel
xs=as.matrix(X1) #variabel spline kn=1 #jumlah titik knot
kn=1 #1 titik knot
ps=ncol(xs) #jumlah variabel spline
pk=ncol(as.matrix(xk)) #jumlah variabel kernel
int.sp=20 #jumlah pembagi titik knot yang diinginkan
alpha=0.05
int.kr=20 #jumlah pembagi titik bandwidth yang diinginkan

#penentuan titik knot
knot=matrix(0,int.sp,ps)
for (i in 1:ps)
{
  knot[,i]=seq(min(xs[,i]),max(xs[,i]),length.out=int.sp)
}
knot=as.matrix(knot[2:(int.sp-1),])
nknot=nrow(knot)

if (kn==1){
  knot=as.matrix(knot)
}else if (kn==2)
{
  #knot2
  nkomb=(nknot*(nknot-1)/2)
  knot2=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-1))
  {
    for (j in (i+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l])
        kk=cbind(kk,a)
      }
    }
  }
}

```

```

    }
    knot2[v,]=kk[1,2:ncol(kk)]
    v=v+1
  }
}
knot=as.matrix(knot2)
nknot=nrow(knot)
}else
{
  #knot3
  nkomb=(nknot*(nknot-1)*(nknot-2)/6)
  knot3=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-2))
  {
    for (j in (i+1):(nknot-1))
    {
      for (k in (j+1):nknot)
      {
        kk=0
        for (l in 1:ps)
        {
          a=cbind(knot[i,l],knot[j,l],knot[k,l])
          kk=cbind(kk,a)
        }
        knot3[v,]=kk[1,2:ncol(kk)]
        v=v+1
      }
    }
  }
  knot=as.matrix(knot3)
  nknot=nrow(knot)
}

#penentuan bandwidth
bw=matrix(0,int.kr,pk)
for (i in 1:pk)
{
  bw[,i]=seq(0,(max(xk[,i])-min(xk[,i])),length.out=int.kr)
}
bw=as.matrix(bw[2:(int.kr-1),])
nband=nrow(bw)

#desain matriks X(k) pada spline
MSE=matrix(0,nband*nknot)
GCV=matrix(0,nband*nknot)
SSE=matrix(0,nband*nknot)
SST=matrix(0,nband*nknot)
R2=matrix(0,nband*nknot)
Normalitas=matrix(0,nband*nknot)
code=matrix(0,nband*nknot,kn*ps+pk)
o=1
for (i in 1:nknot)
{
  for (j in 1:nband)
  {
    #matrik spline
    Z=cbind(1,xs)
    a=1
    for (k in 1:ps)
    {
      for (l in 1:kn)
      {
        Z=cbind(Z,(pmax(0,xs[,k]-knot[i,a])))
      }
    }
  }
}

```

```

        a=a+1
    }
}
sum.v.phi=0
for (k in 1:pk)
{
    v.diag=diag(xk[,k])
    V=m1.nn %*%v.diag
    z=(t(V)-V)/bw[j,k]
    K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
    K.Z=(1/bw[j,k])*K
    W.penyebut=diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
    V.phi=1/n*K.Z/W.penyebut

    #penimbang V(phi).1
    sum.v.phi=sum.v.phi+V.phi #nilai kernel untuk setiap variabel
}
# penimbang kernel gabungan
V.phi=sum.v.phi/pk          #nilai kernel rata-rata

#estimasi parameter
beta=0
C=pinv(t(Z)%*%Z)%*%t(Z)%*%(mi.nn-V.phi)
beta=C%*%Y
A=Z%*%C
B=A+V.phi
yhat=B%*%Y
error=Y-yhat
residual=error
MSE[o]=n^-1*(t(error)%*%error)

SSE=sum((Y-yhat)^2)
SST=sum((Y-mean(Y))^2)
R2[o]=(1-(SSE/SST))

#Normalitas

Normalitas[o]=ks.test(residual,"pnorm",mean(residual),sqrt(var(residual)))$p.value

#Generalized Cross Validation
db=(sum(diag(mi.nn-A-V.phi))/(n))^2
GCV[o]=MSE[i]/(1-db)
code[o,]=c(knot[i,],bw[j,])
o=o+1
}
}
optimum=cbind(code,MSE,GCV,R2,Normalitas)

GCVmin=optimum[order(optimum[, (kn*ps+pk+2)]),]
print(GCVmin[1:1,])

```

**Lampiran 4** *Syntax* R Studi Simulasi Model Regresi Nonparametrik Estimator  
Campuran *Spline Truncated* dan Kernel Gaussian dengan Metode  
*Unbiased Risk (UBR)*

```
#Unbiased Risk
library(pracma)
# Generate some data to test the implementation
n <- 200
eps <- rnorm(n, mean=0, sd = 0.5)
m <- function(x) x^2 * sin(x) #Kernel

f<-function(x) ((sin(pi*x^2))^4)/sin(x^2) #Spline

X1 <- runif(n,0,1)
X2 <- runif(n,0,1)
Y <- f(X1)+m(X2) + eps
xGrid <- seq(-10, 10, l = 500)

par(mfrow=c(1,2))
plot(X1,Y,main="SPLINE")
plot(X2,Y,main="KERNEL")

#matrix
m1.nn=matrix(1, nrow=n, ncol=n) #matriks 1 nxn
m1.n1=matrix(1, nrow=n) #matriks 1 nx1
mi.nn=diag(1,n,n) #matriks identitas nxn

xk=as.matrix(X2) #variabel kernel
xs=as.matrix(X1) #variabel spline kn=1 #jumlah titik knot
kn=1 #1 titik knot
ps=ncol(xs) #jumlah variabel spline
pk.n1=ncol(as.matrix(xk)) #jumlah variabel kernel
int.sp=20 #jumlah pembagi titik knot yang diinginkan
alpha=0.05
int.kr=20 #jumlah pembagi titik bandwidth yang diinginkan

#penentuan titik knot
knot=matrix(0,int.sp,ps)
for (i in 1:ps)
{
  knot[,i]=seq(min(xs[,i]),max(xs[,i]),length.out=int.sp)
}
knot=as.matrix(knot[2:(int.sp-1),])
nknot=nrow(knot)

if (kn==1){
  knot=as.matrix(knot)
}else if (kn==2)
{
  #knot2
  nkomb=(nknot*(nknot-1)/2)
  knot2=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-1))
  {
    for (j in (i+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l])
        kk=cbind(kk,a)
      }
    }
  }
}

```



```

    }
    knot2[v,]=kk[1,2:ncol(kk)]
    v=v+1
  }
}
knot=as.matrix(knot2)
nknot=nrow(knot)
}else
{
  #knot3
  nkomb=(nknot*(nknot-1)*(nknot-2)/6)
  knot3=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-2))
  {
    for (j in (i+1):(nknot-1))
    {
      for (k in (j+1):nknot)
      {
        kk=0
        for (l in 1:ps)
        {
          a=cbind(knot[i,l],knot[j,l],knot[k,l])
          kk=cbind(kk,a)
        }
        knot3[v,]=kk[1,2:ncol(kk)]
        v=v+1
      }
    }
  }
  knot=as.matrix(knot3)
  nknot=nrow(knot)
}

#penentuan bandwidth
bw=matrix(0,int.kr,pk)
for (i in 1:pk)
{
  bw[,i]=seq(0,(max(xk[,i])-min(xk[,i])),length.out=int.kr)
}
bw=as.matrix(bw[2:(int.kr-1),])
nband=nrow(bw)

#desain matriks X(k) pada spline
MSE=matrix(0,nband*nknot)
UBR=matrix(0,nband*nknot)
SSE=matrix(0,nband*nknot)
SST=matrix(0,nband*nknot)
R2=matrix(0,nband*nknot)
Normalitas=matrix(0,nband*nknot)
code=matrix(0,nband*nknot,kn*ps+pk)
o=1
for (i in 1:nknot)
{
  for (j in 1:nband)
  {
    #matrik spline
    Z=cbind(1,xs)
    a=1
    for (k in 1:ps)
    {
      for (l in 1:kn)
      {
        Z=cbind(Z,(pmax(0,xs[,k]-knot[i,a])))

```

```

        a=a+1
    }
}
sum.v.phi=0
for (k in 1:pk)
{
    v.diag=diag(xk[,k])
    V=m1.nn %*%v.diag
    z=(t(V)-V)/bw[j,k]
    K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
    K.Z=(1/bw[j,k])*K
    W.penyebut=diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
    V.phi=1/n*K.Z/W.penyebut

    #penimbang V(phi).1
    sum.v.phi=sum.v.phi+V.phi #nilai kernel untuk setiap variabel
}
# penimbang kernel gabungan
V.phi=sum.v.phi/pk          #nilai kernel rata-rata

#estimasi parameter
beta=0
C=pinv(t(Z)%*%Z)%*%t(Z)%*%(mi.nn-V.phi)
beta=C%*%Y
A=Z%*%C
B=A+V.phi
yhat=B%*%Y
error=Y-yhat
residual=error
MSE[o]=n^-1*(t(error)%*%error)

SSE=sum((Y-yhat)^2)
SST=sum((Y-mean(Y))^2)
R2[o]=(1-(SSE/SST))

#Normalitas
Normalitas[o]=ks.test(residual,"pnorm",mean(residual),sqrt(var(residual)))$p.value

#Unbiased Risk
A1=(mi.nn-A-V.phi)
U1=Norm((A1%*%yhat))
sigmaU=U1/Trace(A1)
U3=sigmaU*Trace(A1^2)/n
U4=sigmaU*Trace(B)/n
UBR[o]=(U1+U3+U4)/n
code[o,]=c(knot[i,],bw[j,])
o=o+1
}
}
optimum=cbind(code,MSE,UBR,R2,Normalitas)

UBRmin=optimum[order(optimum[, (kn*ps+pk+2)]),]
print(UBRmin[1:1,])
}

```

**Lampiran 5** *Syntax R Data Aplikasi pada Model Regresi Nonparametrik Estimator Campuran Spline Truncated dan Kernel Gaussian dengan Metode Cross-Validation (CV)*

```

library(pracma)
data<-read.csv(file.choose(),header=TRUE)
data
y=data[,1] #variabel y
x=as.matrix(data[,2:8])
xk=as.matrix(x[,c(7)]) #variabel kernel
xs=as.matrix(x[,c(1,2,3,4,5,6)]) #variabel spline kn=3 #jumlah titik
knot
kn=3 #3 titik knot
n=length(y) #jumlah pengamatan
pk=ncol(xk) #jumlah variabel kernel
ps=ncol(xs) #jumlah variabel spline

int.kr=40 #jumlah pembagi titik bandwidth yang diinginkan
int.sp=40 #jumlah pembagi titik knot yang diinginkan
alpha=0.05

#matrix
m1.nn=matrix(1, nrow=n, ncol=n) #matriks 1 nxn
m1.n1=matrix(1, nrow=n) #matriks 1 nx1
mi.nn=diag(1,n,n) #matriks identitas nxn

#penentuan titik knot
knot=matrix(0,int.sp,ps)
for (i in 1:ps)
{
  knot[,i]=seq(min(xs[,i]),max(xs[,i]),length.out=int.sp)
}
knot=as.matrix(knot[2:(int.sp-1),])
nknot=nrow(knot)

if (kn==1){
  knot=as.matrix(knot)
}else if (kn==2)
{
  #knot2
  nkomb=(nknot*(nknot-1)/2)
  knot2=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-1))
  {
    for (j in (i+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l])
        kk=cbind(kk,a)
      }
      knot2[v,]=kk[1,2:ncol(kk)]
      v=v+1
    }
  }
  knot=as.matrix(knot2)
  nknot=nrow(knot)
}else
{
  #knot3

```

```

nkomb=(nknot*(nknot-1)*(nknot-2)/6)
knot3=matrix(0,nkomb,kn*ps)
v=1
for (i in 1:(nknot-2))
{
  for (j in (i+1):(nknot-1))
  {
    for (k in (j+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l],knot[k,l])
        kk=cbind(kk,a)
      }
      knot3[v,]=kk[1,2:ncol(kk)]
      v=v+1
    }
  }
}
knot=as.matrix(knot3)
nknot=nrow(knot)
}

#penentuan bandwidth
bw=matrix(0,int.kr,pk)
for (i in 1:pk)
{
  bw[,i]=seq(0,(max(xk[,i])-min(xk[,i])),length.out=int.kr)
}
bw=as.matrix(bw[2:(int.kr-1),])
nband=nrow(bw)

#desain matriks X(k) pada spline
MSE=matrix(0,nband*nknot)
CV=matrix(0,nband*nknot)
code=matrix(0,nband*nknot,kn*ps+pk)
o=1
for (i in 1:nknot)
{
  for (j in 1:nband)
  {
    #matrik spline
    Z=cbind(1,xs)
    a=1
    for (k in 1:ps)
    {
      for (l in 1:kn)
      {
        Z=cbind(Z,(pmax(0,xs[,k]-knot[i,a])))
        a=a+1
      }
    }
    sum.v.phi=0
    for (k in 1:pk)
    {
      v.diag=diag(xk[,k])
      V=m1.nn %*%v.diag
      z=(t(V)-V)/bw[j,k]
      K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
      K.Z=(1/bw[j,k])*K
      W.penyebut=diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
      V.phi=1/n*K.Z/W.penyebut
    }
  }
}

```

```

        #penimbang V(phi).1
        sum.v.phi=sum.v.phi+V.phi #nilai kernel untuk setiap variabel
    }
    # penimbang kernel gabungan
    V.phi=sum.v.phi/pk #nilai kernel rata-rata

    #estimasi parameter
    beta=0
    C=pinv(t(Z)***Z)**t(Z)**(m1.nn-V.phi)
    beta=C***y
    A=Z***C
    B=A+V.phi
    yhat=B***y
    error=y-yhat
    MSE[o]=n^-1*(t(error)***error)

    #Cross Validation
    residual=error
    A3=(rep(1,nrow(B))-diag(B))
    CV[o]=sum((residual/A3)^2)/(n)
    code[o,]=c(knot[i,],bw[j,])
    o=o+1
}
}
optimum=cbind(code,MSE,CV)
optimum
CVmin=optimum[order(optimum[, (kn*ps+pk+2)]),]
CVmin

#mengurutkan nilai CV minimum
knot.opt=CVmin[1,1:(kn*ps)]
band.opt=CVmin[1, (kn+ps+10)]
cv.opt=CVmin[1,ncol(CVmin)]

#validasi nilai CV terkecil
#matrik spline
Z=cbind(1,xs)
a=1
for (k in 1:ps)
{
    for (l in 1:kn)
    {
        Z=cbind(Z, (pmax(0,xs[,k]-knot.opt[a])))
        a=a+1
    }
}

sum.v.phi=0
for (k in 1:pk)
{
    v.diag=diag(xk[,k])
    V=m1.nn ***v.diag
    z=(t(V)-V)/band.opt[k]
    K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
    K.Z=(1/band.opt[k])*K
    W.penyebut=diag(c(1/n*K.Z***m1.n1))***m1.nn
    V.phi=1/n*K.Z/W.penyebut
    #penimbang V(phi).1
    sum.v.phi=sum.v.phi+V.phi #nilai kernel rata-rata setiap variabel
}

# penimbang kernel gabungan
V.phi=sum.v.phi/pk #nilai kernel rata-rata

```

```

#estimasi parameter
beta=0
C=pinv(t(Z)%*%Z)%*%t(Z)%*%(mi.nn-V.phi)
beta=C%*%y
A=Z%*%C
B=A+V.phi
yhat=B%*%y
error=y-yhat
db=matrix(NA,nrow=3)
SS=matrix(NA,nrow=3)
MS=matrix(NA,nrow=3)
PValue=matrix(NA,nrow=3)
deci=matrix(0,nrow=(ps*(kn+1)+1))
Fhitung=matrix(NA,nrow=3)
db[1]=ncol(Z)
db[2]=n-db[1]-1
db[3]=n-1
SS[1]=sum((yhat-mean(y))^2)
SS[2]=sum((y-yhat)^2)
SS[3]=sum((y-mean(y))^2)
MS[1]=SS[1]/db[1]
MS[2]=SS[2]/db[2]
R2=(SS[1]/(SS[1]+SS[2]))*100
MAPE=(sum((abs(y-yhat))/y)/n)*100

#Uji F (Uji Serentak)
Fhitung[1]=MS[1]/MS[2]
PValue[1]=pf(Fhitung[1],db[1],db[2],lower.tail=FALSE)
ANOVA=cbind(db,SS,MS,Fhitung,PValue)
colnames(ANOVA)=c("db","SS","MS","Fhitung","P-Value")
rownames(ANOVA)=c("Regresi","Error","Total")
Ftabel=qf(0.95,db[1],db[2])
if(Fhitung[1]>Ftabel)
{dec='H0 ditolak'
}else
  dec='H0 gagal ditolak'

knot.opt
band.opt
cv.opt
R2
MAPE
ANOVA
Ftabel
dec
beta

residual=y-yhat
plot(y,ylim=c(1.5,13),main="Mixed Estimator Spline & Kernel Gaussian
with CV Method",xlab="Observation",ylab="Percentage of Poor People")
lines(y,col="blue",type="l",lwd=2)
lines(yhat,col="orange",type="o",lwd=2)
legend(21,12,legend=c("Actual","Prediction"),col=c("blue","orange"),lwd
=2)
text(x=41,y=13,label="Coefficient of Determination=83.80%",cex=1)

```

**Lampiran 6** *Syntax R Data Aplikasi pada Model Regresi Nonparametrik Estimator Campuran Spline Truncated dan Kernel Gaussian dengan Metode Generalized Cross-Validation (GCV)*

```

library(pracma)
data<-read.csv(file.choose(),header=TRUE)
data
y=data[,1] #variabel y
x=as.matrix(data[,2:8])
xk=as.matrix(x[,c(7)]) #variabel kernel
xs=as.matrix(x[,c(1,2,3,4,5,6)]) #variabel spline kn=3 #jumlah titik
knot
kn=3 #3 titik knot
n=length(y) #jumlah pengamatan
pk=ncol(xk) #jumlah variabel kernel
ps=ncol(xs) #jumlah variabel spline

int.kr=40 #jumlah pembagi titik bandwidth yang diinginkan
int.sp=40 #jumlah pembagi titik knot yang diinginkan
alpha=0.05

#matrix
m1.nn=matrix(1, nrow=n, ncol=n) #matriks 1 nxn
m1.nl=matrix(1, nrow=n) #matriks 1 nx1
mi.nn=diag(1,n,n) #matriks identitas nxn

#penentuan titik knot
knot=matrix(0,int.sp,ps)
for (i in 1:ps)
{
  knot[,i]=seq(min(xs[,i]),max(xs[,i]),length.out=int.sp)
}
knot=as.matrix(knot[2:(int.sp-1),])
nknot=nrow(knot)

if (kn==1){
  knot=as.matrix(knot)
}else if (kn==2)
{
  #knot2
  nkomb=(nknot*(nknot-1)/2)
  knot2=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-1))
  {
    for (j in (i+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l])
        kk=cbind(kk,a)
      }
      knot2[v,]=kk[1,2:ncol(kk)]
      v=v+1
    }
  }
  knot=as.matrix(knot2)
  nknot=nrow(knot)
}else
{
  #knot3

```

```

nkomb=(nknot*(nknot-1)*(nknot-2)/6)
knot3=matrix(0,nkomb,kn*ps)
v=1
for (i in 1:(nknot-2))
{
  for (j in (i+1):(nknot-1))
  {
    for (k in (j+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l],knot[k,l])
        kk=cbind(kk,a)
      }
      knot3[v,]=kk[1,2:ncol(kk)]
      v=v+1
    }
  }
}
knot=as.matrix(knot3)
nknot=nrow(knot)
}

#penentuan bandwidth
bw=matrix(0,int.kr,pk)
for (i in 1:pk)
{
  bw[,i]=seq(0,(max(xk[,i])-min(xk[,i])),length.out=int.kr)
}
bw=as.matrix(bw[2:(int.kr-1),])
nband=nrow(bw)

#desain matriks X(k) pada spline
MSE=matrix(0,nband*nknot)
GCV=matrix(0,nband*nknot)
code=matrix(0,nband*nknot,kn*ps+pk)
o=1
for (i in 1:nknot)
{
  for (j in 1:nband)
  {
    #matrik spline
    Z=cbind(1,xs)
    a=1
    for (k in 1:ps)
    {
      for (l in 1:kn)
      {
        Z=cbind(Z,(pmax(0,xs[,k]-knot[i,a])))
        a=a+1
      }
    }
    sum.v.phi=0
    for (k in 1:pk)
    {
      v.diag=diag(xk[,k])
      V=m1.nn %*%v.diag
      z=(t(V)-V)/bw[j,k]
      K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
      K.Z=(1/bw[j,k])*K
      W.penyebut=diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
      V.phi=1/n*K.Z/W.penyebut
    }
  }
}

```



```

        #penimbang V(phi).1
        sum.v.phi=sum.v.phi+V.phi #nilai kernel untuk setiap variabel
    }
    # penimbang kernel gabungan
    V.phi=sum.v.phi/pk #nilai kernel rata-rata

    #estimasi parameter
    beta=0
    C=pinv(t(Z)%%Z)%%t(Z)%%(mi.nn-V.phi)
    beta=C%%y
    A=Z%%C
    B=A+V.phi
    yhat=B%%y
    error=y-yhat
    MSE[o]=n^-1*(t(error)%%error)
    db=(sum(diag(mi.nn-A-V.phi))/(n))^2
    GCV[o]=MSE[i]/(db)
    code[o,]=c(knot[i,],bw[j,])
    o=o+1
    }
}
optimum=cbind(code,MSE,GCV)
optimum
GCVmin=optimum[order(optimum[, (kn*ps+pk+2)]),]
GCVmin

#mengurutkan nilai GCV minimum
knot.opt=GCVmin[1,1:(kn*ps)]
band.opt=GCVmin[1,(kn+ps+10)]
gcv.opt=GCVmin[1,ncol(GCVmin)]
#validasi nilai GCV terkecil
#matrik spline
Z=cbind(1,xs)
a=1
for (k in 1:ps)
{
    for (l in 1:kn)
    {
        Z=cbind(Z,(pmax(0,xs[,k]-knot.opt[a])))
        a=a+1
    }
}
sum.v.phi=0
for (k in 1:pk)
{
    v.diag=diag(xk[,k])
    V=m1.nn %%v.diag
    z=(t(V)-V)/band.opt[k]
    K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
    K.Z=(1/band.opt[k])*K
    W.penyebut=diag(c(1/n*K.Z%%m1.n1))%%m1.nn
    V.phi=1/n*K.Z/W.penyebut
    #penimbang V(phi).1
    sum.v.phi=sum.v.phi+V.phi #nilai kernel rata-rata setiap variabel
}

# penimbang kernel gabungan
V.phi=sum.v.phi/pk #nilai kernel rata-rata

#estimasi parameter
beta=0
C=pinv(t(Z)%%Z)%%t(Z)%%(mi.nn-V.phi)
beta=C%%y

```

```

A=Z%*%C
B=A+V.phi
yhat=B%*%y
error=y-yhat
db=matrix(NA,nrow=3)
SS=matrix(NA,nrow=3)
MS=matrix(NA,nrow=3)
PValue=matrix(NA,nrow=3)
deci=matrix(0,nrow=(ps*(kn+1)+1))
Fhitung=matrix(NA,nrow=3)
db[1]=ncol(Z)
db[2]=n-db[1]-1
db[3]=n-1
SS[1]=sum((yhat-mean(y))^2)
SS[2]=sum((y-yhat)^2)
SS[3]=sum((y-mean(y))^2)
MS[1]=SS[1]/db[1]
MS[2]=SS[2]/db[2]
R2=(SS[1]/(SS[1]+SS[2]))*100
MAPE=(sum(abs(y-yhat))/y)/n*100

#Uji F (Uji Serentak)
Fhitung[1]=MS[1]/MS[2]
PValue[1]=pf(Fhitung[1],db[1],db[2],lower.tail=FALSE)
ANOVA=cbind(db,SS,MS,Fhitung,PValue)
colnames(ANOVA)=c("db","SS","MS","Fhitung","P-Value")
rownames(ANOVA)=c("Regresi","Error","Total")
Ftabel=qf(0.95,db[1],db[2])
if (Fhitung[1]>Ftabel)
{dec='H0 ditolak'
}else
  dec='H0 gagal ditolak'

knot.opt
band.opt
gcv.opt
R2
MAPE
ANOVA
Ftabel
dec
beta

residual=y-yhat
y
yhat
cbind(y,yhat)
plot(y,ylim=c(1.5,13),main="Mixed Estimator Spline & Kernel Gaussian
with GCV Method",xlab="Observation",ylab="Percentage of Poor People")
lines(y,col="blue",type="l",lwd=2)
lines(yhat,col="orange",type="o",lwd=2)
legend(21,12,legend=c("Actual","Prediction"),col=c("blue","orange"),lwd
=2)
text(x=41,y=13,label="Coefficient of Determination=90.52%",cex=1)

```

**Lampiran 7** *Syntax R* Data Aplikasi pada Model Regresi Nonparametrik  
Estimator Campuran *Spline Truncated* dan Kernel Gaussian dengan  
Metode *Unbiased Risk* (UBR)

```

library(pracma)
data<-read.csv(file.choose(),header=TRUE)
data
y=data[,1] #variabel y
x=as.matrix(data[,2:8])
xk=as.matrix(x[,c(7)]) #variabel kernel
xs=as.matrix(x[,c(1,2,3,4,5,6)]) #variabel spline kn=3 #jumlah titik
knot
kn=3 #3 titik knot
n=length(y) #jumlah pengamatan
pk=ncol(xk) #jumlah variabel kernel
ps=ncol(xs) #jumlah variabel spline

int.kr=40 #jumlah pembagi titik bandwidth yang diinginkan
int.sp=40 #jumlah pembagi titik knot yang diinginkan
alpha=0.05

#matrix
m1.nn=matrix(1, nrow=n, ncol=n) #matriks 1 nxn
m1.nl=matrix(1, nrow=n) #matriks 1 nx1
mi.nn=diag(1,n,n) #matriks identitas nxn

#penentuan titik knot
knot=matrix(0,int.sp,ps)
for (i in 1:ps)
{
  knot[,i]=seq(min(xs[,i]),max(xs[,i]),length.out=int.sp)
}
knot=as.matrix(knot[2:(int.sp-1),])
nknot=nrow(knot)

if (kn==1){
  knot=as.matrix(knot)
}else if (kn==2)
{
  #knot2
  nkomb=(nknot*(nknot-1)/2)
  knot2=matrix(0,nkomb,kn*ps)
  v=1
  for (i in 1:(nknot-1))
  {
    for (j in (i+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l])
        kk=cbind(kk,a)
      }
      knot2[v,]=kk[1,2:ncol(kk)]
      v=v+1
    }
  }
  knot=as.matrix(knot2)
  nknot=nrow(knot)
}else
{
  #knot3

```

```

nkomb=(nknot*(nknot-1)*(nknot-2)/6)
knot3=matrix(0,nkomb,kn*ps)
v=1
for (i in 1:(nknot-2))
{
  for (j in (i+1):(nknot-1))
  {
    for (k in (j+1):nknot)
    {
      kk=0
      for (l in 1:ps)
      {
        a=cbind(knot[i,l],knot[j,l],knot[k,l])
        kk=cbind(kk,a)
      }
      knot3[v,]=kk[1,2:ncol(kk)]
      v=v+1
    }
  }
}
knot=as.matrix(knot3)
nknot=nrow(knot)
}

#penentuan bandwidth
bw=matrix(0,int.kr,pk)
for (i in 1:pk)
{
  bw[,i]=seq(0,(max(xk[,i])-min(xk[,i])),length.out=int.kr)
}
bw=as.matrix(bw[2:(int.kr-1),])
nband=nrow(bw)

#desain matriks X(k) pada spline
MSE=matrix(0,nband*nknot)
UBR=matrix(0,nband*nknot)
code=matrix(0,nband*nknot,kn*ps+pk)
o=1
for (i in 1:nknot)
{
  for (j in 1:nband)
  {
    #matrik spline
    Z=cbind(1,xs)
    a=1
    for (k in 1:ps)
    {
      for (l in 1:kn)
      {
        Z=cbind(Z,(pmax(0,xs[,k]-knot[i,a])))
        a=a+1
      }
    }
    sum.v.phi=0
    for (k in 1:pk)
    {
      v.diag=diag(xk[,k])
      V=m1.nn %*%v.diag
      z=(t(V)-V)/bw[j,k]
      K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
      K.Z=(1/bw[j,k])*K
      W.penyebut=diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
      V.phi=1/n*K.Z/W.penyebut
    }
  }
}

```

```

#penimbang V(phi).1
sum.v.phi=sum.v.phi+V.phi #nilai kernel untuk setiap variabel
}
# penimbang kernel gabungan
V.phi=sum.v.phi/pk #nilai kernel rata-rata

#estimasi parameter
beta=0
C=pinv(t(Z)%%Z)%%t(Z)%(mi.nn-V.phi)
beta=C%%y
A=Z%%C
B=A+V.phi
yhat=B%%y
error=y-yhat
MSE[o]=n^-1*(t(error)%%error)

A1=(mi.nn-A-V.phi)
U1=Norm((A1%%yhat))
sigmaU=U1/Trace(A1)
U3=sigmaU*Trace(A1^2)/n
U4=sigmaU*Trace(B)/n
UBR[o]=(U1+U3+U4)/n
code[o,]=c(knot[i,],bw[j,])
o=o+1
}
}
optimum=cbind(code,MSE,UBR)

UBRmin=optimum[order(optimum[, (kn*ps+pk+2)]),]
UBRmin

#mengurutkan nilai UBR minimum
knot.opt=UBRmin[1,1:(kn*ps)]
band.opt=UBRmin[1,(kn+ps+10)]
ubr.opt=UBRmin[1,ncol(UBRmin)]

#validasi nilai UBR terkecil
#matrik spline
Z=cbind(1,xs)
a=1
for (k in 1:ps)
{
for (l in 1:kn)
{
Z=cbind(Z,(pmax(0,xs[,k]-knot.opt[a])))
a=a+1
}
}
}
sum.v.phi=0
for (k in 1:pk)
{
v.diag=diag(xk[,k])
V=m1.nn %%v.diag
z=(t(V)-V)/band.opt[k]
K=1/sqrt(2*pi)*exp(-1/2*z^2) #fungsi kernel gaussian
K.Z=(1/band.opt[k])*K
W.penyebut=diag(c(1/n*K.Z%%m1.n1))%%m1.nn
V.phi=1/n*K.Z/W.penyebut
#penimbang V(phi).1
sum.v.phi=sum.v.phi+V.phi #nilai kernel rata-rata setiap variabel
}

# penimbang kernel gabungan
V.phi=sum.v.phi/pk #nilai kernel rata-rata

```

```

#estimasi parameter
beta=0
C=pinv(t(Z)%*%Z)%*%t(Z)%*%(mi.nn-V.phi)
beta=C%*%y
A=Z%*%C
B=A+V.phi
yhat=B%*%y
error=y-yhat
db=matrix(NA,nrow=3)
SS=matrix(NA,nrow=3)
MS=matrix(NA,nrow=3)
PValue=matrix(NA,nrow=3)
deci=matrix(0,nrow=(ps*(kn+1)+1))
Fhitung=matrix(NA,nrow=3)
db[1]=ncol(Z)
db[2]=n-db[1]-1
db[3]=n-1
SS[1]=sum((yhat-mean(y))^2)
SS[2]=sum((y-yhat)^2)
SS[3]=sum((y-mean(y))^2)
MS[1]=SS[1]/db[1]
MS[2]=SS[2]/db[2]
R2=(SS[1]/(SS[1]+SS[2]))*100
MAPE=((sum(abs(y-yhat))/y)/n)*100

#Uji F (Uji Serentak)
Fhitung[1]=MS[1]/MS[2]
PValue[1]=pf(Fhitung[1],db[1],db[2],lower.tail=FALSE)
ANOVA=cbind(db,SS,MS,Fhitung,PValue)
colnames(ANOVA)=c("db","SS","MS","Fhitung","P-Value")
rownames(ANOVA)=c("Regresi","Error","Total")
Ftabel=qf(0.95,db[1],db[2])
if(Fhitung[1]>Ftabel)
{dec='H0 ditolak'
}else
  dec='H0 gagal ditolak'

knot.opt
band.opt
ubr.opt
R2
MAPE
ANOVA
Ftabel
dec
beta

residual=y-yhat

cbind(y,yhat)
plot(y,ylim=c(1.5,13),main="Mixed Estimator Spline & Kernel Gaussian
with UBR Method",xlab="Observation",ylab="Percentage of Poor People")
lines(y,col="blue",type="l",lwd=2)
lines(yhat,col="orange",type="o",lwd=2)
legend(21,12,legend=c("Actual","Prediction"),col=c("blue","orange"),lwd
=2)
text(x=41,y=13,label="Coefficient of Determination=79.23%",cex=1)

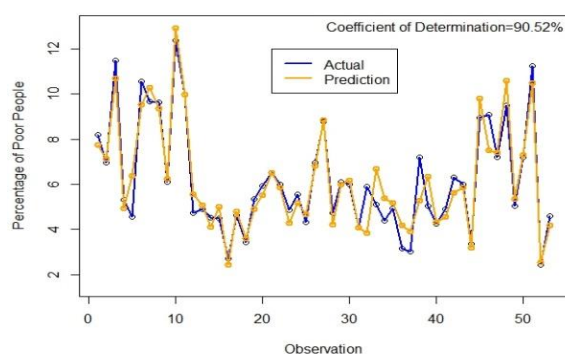
```

**Lampiran 8** Estimasi Variabel Respon ( $\hat{y}$ ) dari Model Regresi Nonparametrik

Estimator Campuran Terbaik (GCV 3 Titik Knot)

Kabupaten/Kota	$y$	$\hat{y}$	Kabupaten/Kota	$y$	$\hat{y}$
Sambas	8,19	7,75	Tana Tidung	4,72	4,20
Bengkayang	6,96	7,13	Nunukan	6,11	5,99
Landak	11,47	10,68	Kota Tarakan	6	6,16
Mempawah	5,32	4,94	Kotawaringin Barat	4,11	4,07
Sanggau	4,57	6,37	Kotawaringin Timur	5,9	3,83
Ketapang	10,54	9,51	Kapuas	5,09	6,67
Sintang	9,65	10,30	Barito Selatan	4,39	5,36
Kapuas Hulu	9,62	9,35	Barito Utara	4,95	5,17
Sekadau	6,11	6,23	Sukamara	3,16	4,19
Melawi	12,38	12,91	Lamandau	3,01	3,91
Kayong Utara	9,98	9,97	Seruyan	7,19	5,27
Kubu Raya	4,74	5,56	Katingan	5,02	6,33
Kota Singkawang	4,91	5,05	Pulang Pisau	4,24	4,30
Tanah Laut	4,51	4,11	Gunung Mas	4,91	4,55
Kota Baru	4,49	5,00	Barito Timur	6,32	5,60
Banjar	2,72	2,45	Murung Raya	6	5,87
Barito Kuala	4,63	4,81	Kota Palangka Raya	3,35	3,20
Tapin	3,41	3,60	Paser	8,95	9,80
Hulu Sungai Selatan	5,33	4,91	Kutai Barat	9,09	7,49
Hulu Sungai Tengah	5,93	5,51	Kutai Kartanegara	7,2	7,40
Hulu Sungai Utara	6,5	6,50	Kutai Timur	9,48	10,60
Tabalong	6,01	5,87	Berau	5,04	5,35
Tanah Bumbu	4,85	4,29	Penajam Paser Utara	7,18	7,31
Balangan	5,55	5,17	Mahakam Ulu	11,25	10,49
Kota Banjar Baru	4,3	4,66	Kota Balikpapan	2,42	2,52
Malinau	6,96	6,82	Kota Samarinda	4,59	4,18
Bulungan	8,78	8,84			

Mixed Estimator Spline & Kernel Gaussian with GCV Method



## BIOGRAFI PENULIS



**Andrea Tri Rian Dani**, lahir pada tanggal 16 Januari 1998 di Samarinda, Kalimantan Timur. Anak dari tiga bersaudara pasangan Bapak H. Mariadi dan Ibu Hj. Sanatun. Memulai pendidikan formal pada Tahun 2003 di Sekolah Dasar Negeri 04 Bukuan Samarinda dan lulus pada Tahun 2009. Menempuh pendidikan selanjutnya di Sekolah Menengah Pertama Negeri 20 Bukuan Samarinda dan lulus pada Tahun 2012, kemudian melanjutkan pendidikan di Sekolah Menengah Atas Negeri 06 Samarinda dan lulus pada Tahun 2015.

Pendidikan perguruan tinggi Sarjana (S1) dimulai pada Tahun 2015 di Universitas Mulawarman, Fakultas Matematika dan Ilmu Pengetahuan Alam, Jurusan Matematika, Program Studi Statistika melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), dan lulus pada Maret 2019. Selama periode Januari hingga Juni 2019, penulis bekerja sebagai Enumerator SPH Bank Indonesia KPw Provinsi Kalimantan Timur. Pada Juli 2019, penulis mendapat kesempatan untuk melanjutkan pendidikan Program Pascasarjana (S2) di Departemen Statistika, Fakultas Sains dan Analitika Data (FSAD) Institut Teknologi Sepuluh Nopember (ITS) Surabaya dengan Beasiswa “Kaltim Tuntas” Provinsi Kalimantan Timur. Pembaca yang ingin memberikan kritik, saran dan diskusi mengenai penelitian ini dapat menghubungi penulis melalui Email: [andreatriandani98@gmail.com](mailto:andreatriandani98@gmail.com).