



---

TESIS - SS142501

# **REGRESI CAMPURAN NONPARAMETRIK SPLINE LINIER TRUNCATED DAN FUNGSI KERNEL UNTUK PEMODELAN DATA KEMISKINAN DI PROVINSI PAPUA**

RORY  
NRP. 1314 201 711

DOSEN PEMBIMBING  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Dr. Wahyu Wibowo, S.Si., M.Si.

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2016



---

THESIS - SS142501

# **NONPARAMETRIC MIXED REGRESSION OF TRUNCATED LINEAR SPLINE AND KERNEL FUNCTION FOR POVERTY DATA MODELING IN PAPUA PROVINCE**

RORY  
NRP. 1314 201 711

SUPERVISOR  
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Dr. Wahyu Wibowo, S.Si., M.Si.

MAGISTER PROGRAM  
DEPARTEMENT OF STATISTICS  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2016

**REGRESI CAMPURAN NONPARAMETRIK SPLINE LINIER  
TRUNCATED DAN FUNGSI KERNEL UNTUK PEMODELAN  
DATA KEMISKINAN DI PROVINSI PAPUA**

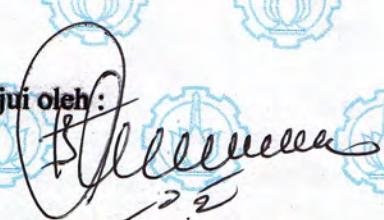
Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Sains (M.Si.)

di  
Institut Teknologi Sepuluh Nopember  
oleh :

**R O R Y**  
**NRP. 1314 201 711**

Tanggal Ujian : 26 Januari 2016  
Periode Wisuda : Maret 2016

Disetujui oleh :



1. Prof. Dr. Drs. I Nyoman Budiantara, M.Si. (Pembimbing I)  
NIP. 19650603 198903 1 003

2. Dr. Wahyu Wibowo, S.Si., M.Si. (Pembimbing II)  
NIP. 19740328 199802 1 001

3. Dr. Dra. Ismaili Zain, M.Si. (Pengaji)  
NIP. 19600525 198803 2 001

4. Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si. (Pengaji)  
NIP. 19831204 200812 1 002

5. Dr. Vera Lisma, S.Si., M.Phil. (Pengaji)  
NIP. 19681107 199403 2 002



Direktur Program Pascasarjana,

Prof. Ir. Djauhar Manfaat, M.Sc, Ph.D.  
NIP. 19601202 198701 1 001

PROGRAM PASCASARJANA



# REGRESI CAMPURAN NONPARAMETRIK SPLINE LINIER TRUNCATED DAN FUNGSI KERNEL UNTUK PEMODELAN DATA KEMISKINAN DI PROVINSI PAPUA

Nama Mahasiswa : Rory  
NRP : 1314 201 711  
Pembimbing : Prof. Dr. Drs. I Nyoman Budiantara, M.Si.  
Co-Pembimbing : Dr. Wahyu Wibowo, S.Si., M.Si.

## ABSTRAK

Model regresi campuran nonparametrik  $y_i = f(u_i, \tilde{v}_i) + \varepsilon_i, i = 1, 2, \dots, n$ ,  $\tilde{v}_i = (v_{1i}, v_{2i}, \dots, v_{mi})^T$  memiliki kurva regresi bersifat aditif  $f(u_i, \tilde{v}_i) = g(u_i) + \sum_{j=1}^m h_j(v_{ji})$ . Komponen  $g(u_i)$  dihampiri dengan spline linier truncated, sedangkan komponen  $h_j(v_{ji})$  dihampiri dengan kernel Nadaraya-Watson. Error random  $\varepsilon_i$  mengikuti distribusi normal  $N(0, \sigma^2)$ . Tujuan dari penelitian ini adalah melakukan kajian mengenai estimator kurva regresi campuran nonparametrik spline dan kernel  $f(u_i, \tilde{v}_i)$  dan mengaplikasikannya pada data kemiskinan di Provinsi Papua. Hasil kajian menunjukkan bahwa estimator kurva regresi spline  $\hat{g}(u)$  adalah  $\hat{\hat{g}}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi})\tilde{y}$  dan estimator kurva regresi kernel  $\sum_{j=1}^m \tilde{h}_j(v_j)$  adalah  $\sum_{j=1}^m \hat{\tilde{h}}_j(v_j) = \mathbf{V}(\tilde{\phi})\tilde{y}$ . Selanjutnya, estimator kurva regresi campuran nonparametrik spline dan kernel  $\tilde{f}(u, \tilde{v})$  adalah  $\hat{\tilde{f}}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{Z}(\tilde{\xi}, \tilde{\phi})\tilde{y}$ , dimana  $\mathbf{Z}(\tilde{\xi}, \tilde{\phi}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})$ . Matriks  $\mathbf{S}(\tilde{\xi}, \tilde{\phi})$ ,  $\mathbf{V}(\tilde{\phi})$  dan  $\mathbf{Z}(\tilde{\xi}, \tilde{\phi})$  tergantung pada lokasi titik-titik knot  $\tilde{\xi}$  dan bandwidth  $\tilde{\phi}$ . Estimator-estimator tersebut adalah estimator bias, namun masih kelas estimator linier. Model regresi campuran nonparametrik terbaik adalah model yang menggunakan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum yang diperoleh dengan meminimumkan fungsi *Generalized Cross Validation* (GCV). Pemilihan lokasi titik-titik knot dan bandwidth dilakukan secara simultan. Model regresi campuran nonparametrik spline dan kernel diterapkan pada data kemiskinan di Provinsi Papua, dimana sebagai variabel responnya adalah persentase penduduk miskin ( $y$ ), variabel prediktor yang mengikuti kurva regresi spline adalah PDRB perkapita ( $u$ ), dan variabel-variabel prediktor yang mengikuti kurva regresi kernel adalah gini ratio ( $v_1$ ), rata-rata lama sekolah ( $v_2$ ), tingkat pengangguran terbuka ( $v_3$ ) dan laju pertumbuhan ekonomi ( $v_4$ ). Model terbaik diperoleh ketika model menggunakan 3 titik knot. Estimasi model memberikan  $R^2$  sebesar 92,02%. Model dapat digunakan untuk skenario kebijakan.

Kata kunci: kernel nadaraya-watson, regresi campuran nonparametrik, regresi nonparametrik aditif, spline linier truncated

# NONPARAMETRIC MIXED REGRESSION OF TRUNCATED LINEAR SPLINE AND KERNEL FUNCTION FOR POVERTY DATA MODELING IN PAPUA PROVINCE

Name

: Rory

NRP

: 1314 201 711

Supervisor

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

Co-Supervisor

: Dr. Wahyu Wibowo, S.Si., M.Si.

## ABSTRACT

Nonparametric mixed regression model  $y_i = f(u_i, \tilde{v}_i) + \varepsilon_i$ ,  $i = 1, 2, \dots, n$ ,  $\tilde{v}_i = (v_{1i}, v_{2i}, \dots, v_{mi})^T$  has additive regression curve  $f(u_i, \tilde{v}_i) = g(u_i) + \sum_{j=1}^m h_j(v_{ji})$ . Component of  $g(u_i)$  is approached by truncated linear spline regression curve, while components of  $h_j(v_{ji})$  is approached by Nadaraya-Watson kernel regression curve. Random error  $\varepsilon_i$  follows independent normal distribution normal  $N(0, \sigma^2)$ . The aim of this thesis is to perform a study on estimator of nonparametric mixed regression curve of spline and kernel  $f(u_i, \tilde{v}_i)$  and to apply it to poverty data in Papua Province. The results of study show that estimator of spline regression curve  $\tilde{g}(u)$  is  $\hat{\tilde{g}}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi})\tilde{y}$  and estimator of kernel regression curve  $\sum_{j=1}^m h_j(v_j)$  is  $\sum_{j=1}^m \hat{h}_j(v_j) = \mathbf{V}(\tilde{\phi})\tilde{y}$ . Furthermore, estimator of nonparametric mixed regression curve of spline and kernel  $\tilde{f}(u, \tilde{v})$  is  $\hat{\tilde{f}}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{Z}(\tilde{\xi}, \tilde{\phi})\tilde{y}$ , where  $\mathbf{Z}(\tilde{\xi}, \tilde{\phi}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})$ . The estimators are biased estimator, but still the class of linear estimator. Matrix  $\mathbf{S}(\tilde{\xi}, \tilde{\phi})$ ,  $\mathbf{V}(\tilde{\phi})$  and  $\mathbf{Z}(\tilde{\xi}, \tilde{\phi})$  depend on knot points location  $\tilde{\xi}$  and bandwidths  $\tilde{\phi}$ . The best nonparametric mixed regression model of spline and kernel is the model that uses the optimum of number of knots point, knot points location and bandwidths obtained from Generalized Cross Validation (GCV) method. The choice of knot points location and bandwidths is conducted simultaneously. Nonparametric mixed regression model of spline and kernel is applied to poverty data in Papua Province, in which the response variable is percentage of poor people ( $y$ ), the predictor variable that follows spline regression curve is GDRP per capita ( $u$ ), and the predictor variables that follow kernel curve regression is gini ratio ( $v_1$ ), mean years of schooling ( $v_2$ ), unemployment rate ( $v_3$ ) and economic growth ( $v_4$ ). The best model is obtained when the model has 3 knot points. Estimation model results  $R^2$  of 92.02%. The model can be used for policy scenario.

Key words: additive nonparametric regression, kernel nadaraya-watson, nonparametric mixed regression, truncated linear spline

## KATA PENGANTAR

Alhamdulillah, puji syukur penulis panjatkan kehadiran Allah SWT, atas segala limpahan rahmat dan kemurahan-Nya penulis dapat menyelesaikan tesis dengan judul “*Regresi Campuran Nonparametrik Spline Linier Truncated dan Fungsi Kernel untuk Pemodelan Data Kemiskinan di Provinsi Papua*”. Penyelesaian Tesis ini tak lepas dari peranan, dukungan dan motivasi berbagai pihak. Oleh karena itu pada kesempatan ini penulis ingin menyampaikan terima kasih yang sedalam-dalamnya secara khusus kepada kedua orang tua yang sangat penulis cintai dan hormati, Bapak Amziur dan Ibu Ermeli. Terima kasih atas segala cinta, pendidikan, doa dan motivasi yang tiada henti. Selanjutnya penulis ingin menyampaikan rasa terimakasih kepada

1. Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si. selaku dosen pembimbing pertama atas motivasi dan kesabarannya dalam mengarahkan penulis dalam menyelesaikan tesis ini,
2. Bapak Dr. Wahyu Wibowo, S.Si., M.Si. selaku dosen pembimbing kedua atas semua petunjuk, saran dan arahannya,
3. Ibu Dr. Dra. Ismaini Zain, M.Si., Bapak Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si., dan Ibu Dr. Vera Lisna, S.Si., M.Phil. selaku dosen pengujii yang telah memberikan banyak saran dan masukan terhadap tesis ini.
4. Bapak Dr. Suhartono, S.Si., M.Sc. selaku ketua Jurusan Statistika FMIPA ITS, atas segala kemudahan urusan akademis dan fasilitas yang menunjang di Jurusan Statistika FMIPA ITS Surabaya.
5. Bapak Ibu Dosen selaku Pengajar di Jurusan Statistika atas pembekalan ilmu kepada penulis selama penulis menempuh pendidikan di Program Studi Magister Jurusan Statistika FMIPA ITS Surabaya. Terimakasih juga kepada Seluruh Staf Pegawai di Jurusan Statistika FMIPA ITS atas pelayanannya yang tulus dan ramah.
6. Pimpinan Badan Pusat Statistik (BPS) RI beserta seluruh jajarannya, atas kesempatan yang diberikan kepada penulis untuk melanjutkan pendidikan S2.

7. Pimpinan BPS Provinsi Jambi dan PUSDIKLAT BPS yang telah memberikan kesempatan dan motivasi kepada penulis untuk menempuh pendidikan magister di Jurusan Statistika FMIPA ITS.
8. Pimpinan BPS Kabupaten Tanjung Jabung Barat atas kepercayaan dan ijin yang telah diberikan serta Seluruh Pegawai BPS Kabupaten Tanjung Jabung Barat atas suka duka, kenangan dan pengalaman yang diberikan.
9. Saudariku Uni Elfina Rozita, Uda Ridsalman dan keponakanku Muhammad Alif Salman atas motivasi dan semangatnya.
10. Nani Sundari, yang telah memberikan motivasi, semangat dan doanya selama menyelesaikan tesis ini.
11. Saudara senasib sepenanggungan Magister BPS angkatan 8, Pak Komting Mas Ali, Aan, Mpih, Yanti, Mas Mur, Mb Nike, Widi, Mas Henri, Maul, Mas Duto, Mas Arip, Daeng Zablin, Dian, Yani, Fatih, Mb Afni, Santi, Vivin, Mb Anita (diurutkan sesuai absen, semoga selalu ingat) atas kerjasama dan kenangan suka dukanya selama menyelesaikan studi.
12. Teman-teman S2 TB BPS angkatan 9 dan Mas-mas S3 TB BPS Jurusan Statistika FMIPA ITS atas dukungan dan bantuannya.
13. Teman-teman S2 Statistika reguler angkatan 2014 yang ikut juga berjuang bersama angkatan 8 TB BPS di Jurusan Statistika FMIPA ITS.
14. Semua pihak yang telah membantu selama penulis menyelesaikan studi, yang tidak dapat penulis sebutkan satu persatu.

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

Penulis

Surabaya, Januari 2016

## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
ABSTRAK .....	iii
ABSTRACT .....	v
KATA PENGANTAR .....	vii
DAFTAR ISI .....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xiii
DAFTAR LAMPIRAN .....	xv
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	4
1.3 Tujuan Penelitian .....	5
1.4 Manfaat Penelitian .....	5
1.5 Batasan Penelitian .....	6
BAB 2 KAJIAN PUSTAKA .....	7
2.1 Analisis Regresi .....	7
2.2 Regresi Parametrik, Nonparametrik dan Semiparametrik .....	7
2.3 Regresi Parametrik Linier Berganda .....	9
2.4 Regresi Nonparametrik Spline Truncated .....	10
2.5 Regresi Nonparametrik Kernel .....	14
2.6 Regresi Campuran Nonparametrik .....	18
2.7 Optimasi nlminb .....	20
2.8 Uji Normalitas Kolmogorov-Smirnov .....	21
2.9 Kriteria Kebaikan Model .....	21
2.10 Tinjauan Kemiskinan .....	22
2.10.1 Konsep dan Pengukuran Kemiskinan .....	22
2.10.2 Faktor-faktor yang Mempengaruhi Kemiskinan .....	23
BAB 3 METODE PENELITIAN .....	27
3.1 Sumber Data .....	27

3.2 Variabel Penelitian.....	27
3.3 Tahapan Penelitian.....	29
<b>BAB 4 HASIL DAN PEMBAHASAN .....</b>	<b>35</b>
4.1 Model Regresi Campuran Nonparametrik Spline dan Kernel .....	35
4.2 Estimasi Kurva Regresi .....	37
4.3 Sifat Estimator Kurva Regresi .....	43
4.4 Pemilihan Titik-titik Knot dan Bandwidth Secara Simultan .....	45
4.5 Aplikasi pada Data Kemiskinan .....	46
4.5.1 Eksplorasi Data.....	46
4.5.2 Model Regresi Campuran Nonparametrik .....	48
4.5.3 Fungsi Kernel, Titik Knot dan Bandwidth Optimum.....	49
4.5.3.1 Model dengan Satu Titik Knot .....	49
4.5.3.2 Model dengan Dua Titik Knot.....	50
4.5.3.3 Model dengan Tiga Titik Knot.....	51
4.5.3.4 GCV Minimum dan Banyaknya Titik Knot Optimum....	52
4.5.4 Estimasi Model Regresi.....	53
4.5.5 Uji Normalitas Residual .....	54
4.5.6 Interpretasi Model .....	55
4.5.7 Simulasi Data untuk Prediksi .....	58
<b>BAB 5 KESIMPULAN .....</b>	<b>61</b>
5.1 Kesimpulan .....	61
5.2 Saran .....	62
<b>DAFTAR PUSTAKA .....</b>	<b>63</b>
<b>LAMPIRAN .....</b>	<b>67</b>
<b>BIOGRAFI PENULIS .....</b>	<b>81</b>

## DAFTAR GAMBAR

Gambar 3.1 Peta Provinsi Papua.....	27
Gambar 3.2 Diagram Alir Tahap Analisis untuk Tujuan Pertama dan Kedua.....	32
Gambar 3.3 Diagram Alir Tahap Analisis untuk Tujuan Ketiga .....	33
Gambar 4.1 Diagram Pencar Variabel Respon dan Prediktor.....	47
Gambar 4.2 <i>Probability Plot Residual</i> .....	54

## DAFTAR TABEL

Tabel 4.1 Statistik Deskriptif Variabel Respon dan Prediktor .....	48
Tabel 4.2 GCV pada Model dengan Satu Titik Knot .....	50
Tabel 4.3 GCV pada Model dengan Dua Titik Knot.....	51
Tabel 4.4 GCV pada Model dengan Tiga Titik Knot.....	52
Tabel 4.5 Perbandingan nilai GCV Minimum .....	52
Tabel 4.6 Skenario Data untuk Prediksi Persentase Penduduk Miskin .....	59

## **DAFTAR LAMPIRAN**

Lampiran 1. Data dan Format Data yang Digunakan .....	67
Lampiran 2. Fungsi nlminb pada Program R .....	68
Lampiran 3. Prosedur Pemanggilan Fungsi nlminb .....	70
Lampiran 4. Syntax R GCV Model dengan Satu Titik Knot .....	72
Lampiran 5. Syntax R GCV Model dengan Dua Titik Knot.....	74
Lampiran 6. Syntax R GCV Model dengan Tiga Titik Knot.....	76
Lampiran 7. Syntax R Estimasi Parameter Model dengan Tiga Titik Knot .....	78
Lampiran 8. Syntax R untuk Menghitung Estimasi $y_i$ dan Residual.....	79
Lampiran 9. Syntax R Prediksi Persentase Penduduk Miskin dari Model .....	80

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi adalah salah satu metode statistika yang sering digunakan oleh para peneliti di berbagai bidang penelitian. Analisis ini digunakan untuk mengetahui pola hubungan dua atau lebih variabel dalam bentuk fungsional. Dalam hubungan fungsional tersebut, masing-masing variabel dikelompokkan ke dalam variabel respon dan variabel prediktor. Identifikasi awal adanya pola hubungan dapat dilakukan dengan memanfaatkan pengalaman masa lalu atau menggunakan diagram pencar (*scatter plot*). Jika bentuk pola hubungan fungsionalnya diketahui, maka model regresi yang digunakan adalah model regresi parametrik. Sebaliknya, jika bentuk pola hubungan fungsionalnya tidak diketahui, maka model regresi yang digunakan adalah model regresi nonparametrik (Eubank, 1999).

Seringkali dalam prakteknya, bentuk pola hubungan fungsional antara variabel respon dan variabel prediktor tidak diketahui. Pada kondisi ini tentu model regresi parametrik kurang cocok digunakan. Oleh karena itu, seiring dengan makin berkembangnya teknologi di bidang komputasi, model regresi nonparametrik yang umumnya membutuhkan tingkat penghitungan yang cukup sulit sudah semakin banyak digunakan. Model regresi nonparametrik ini sangat baik digunakan untuk pola data yang tidak diketahui karena memiliki fleksibilitas yang tinggi, dimana data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh subyektifitas peneliti (Eubank, 1999).

Ada banyak estimator kurva regresi nonparametrik yang telah dikembangkan oleh para peneliti, diantaranya spline (Reinsch, 1967; Silverman, 1985; Wahba, 1990; Eubank, 1999; Liang, 2006; Lin & Zhang, 2006; Islamiyat & Budiantara, 2007; Budiantara, 2009), kernel (Nadaraya, 1989; Gasser & Muller, 1979; Hardle, 1994; Wand & Jones, 1995; Eubank, 1999; You & Chen, 2007; Kayri & Zirbhoglu, 2009; Klemela, 2014) dan deret fourier (De Jong, 1977; Bilodeau, 1992; Morton & Silverberg, 1999; Tripena & Budiantara, 2007; Sudiarsa, Budiantara, Suhartono & Purnami, 2015).

Regresi spline merupakan salah satu model regresi yang mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik (Eubank, 1999). Kurva regresi spline memiliki kemampuan yang sangat baik dalam menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Budiantara, 2009). Dengan bantuan titik-titik knot, kurva spline dapat mengatasi permasalahan pola data yang menunjukkan naik/turun yang tajam sehingga dapat dihasilkan kurva yang relatif mulus (Eubank, 1999). Bentuk kurva regresi spline sangat dipengaruhi oleh banyaknya titik knot dan lokasi titik-titik knot.

Estimator kernel adalah pengembangan dari estimator histogram. Estimator ini merupakan estimator linier yang mirip dengan estimator regresi nonparametrik yang lain, perbedaannya hanya karena estimator kernel lebih khusus dalam penggunaan metode bandwidth (Eubank, 1999). Kelebihan dari estimator kernel adalah memiliki kemampuan yang baik dalam memodelkan data yang tidak mempunyai pola tertentu (Härdle, 1994). Selain itu, estimator kernel lebih fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara & Mulianah, 2007). Dari segi komputasinya, metode kernel lebih mudah dilakukan dan mudah diimplementasikan (Klemela, 2014).

Model regresi nonparametrik dapat diterapkan di berbagai bidang keilmuan, salah satunya adalah bidang sosial. Pada bidang ini, model regresi nonparametrik dapat diterapkan untuk menganalisis masalah kemiskinan. Masalah kemiskinan merupakan masalah yang sangat serius yang dihadapi oleh negara-negara di dunia terutama negara-negara berkembang termasuk Indonesia. Pengentasan kemiskinan yang menjadi tujuan utama *Millennium Development Goals* (MDGs) yang ditetapkan oleh Perserikatan Bangsa-Bangsa (PBB) kembali menjadi tujuan utama dalam *Sustainable Development Goals* (SDGs) untuk tahun 2016–2030. Berdasarkan data BPS (2014), pada September 2013 persentase penduduk miskin di Indonesia sebesar 11,47 persen. Dari 33 provinsi di Indonesia, Provinsi Papua menjadi provinsi yang persentase penduduk miskinnya tertinggi di Indonesia yaitu sebesar 31,52 persen. Hal ini tentu wajib menjadi perhatian pemerintah pusat maupun pemerintah daerah untuk menekan angka kemiskinan di provinsi tersebut.

Masalah kemiskinan merupakan masalah yang komplek dan kronis. Oleh karena itu, menganalisis masalah kemiskinan membutuhkan metode analisis statistik yang tepat, serta melibatkan komponen permasalahan yang tepat pula. Model regresi nonparametrik yang didukung dengan variabel-variabel yang tepat diharapkan bisa menjadi model alternatif yang mampu menganalisis permasalahan kemiskinan.

Penelitian mengenai kemiskinan menggunakan model regresi nonparametrik sebelumnya pernah dilakukan oleh Budiantara, Ratna, Zain, & Wibowo (2012) yang memodelkan kemiskinan di Indonesia menggunakan pendekatan spline truncated. Selanjutnya Merdekawati & Budiantara (2013) memodelkan kemiskinan di Provinsi Jawa Tengah menggunakan pendekatan spline truncated multivariabel. Budiantara, Diana, Purhadi & Darmesto (2011) menganalisis pola hubungan kemiskinan dan pengangguran di Indonesia dengan menggunakan pendekatan spline bayesian. Selanjutnya, Ratnasari, Budiantara, Zain, Ratna & Mirah (2015) menganalisis kemiskinan di Provinsi Papua dengan membandingkan spline truncated dan deret fourier. Muller (1997) menganalisis kemiskinan di Rwanda menggunakan kernel, Jalan & Ravallion (2003) menganalisis kemiskinan di Argentina menggunakan kernel, Sala-i-Martin (2006) menganalisis kemiskinan negara-negara di dunia menggunakan kernel, Minoiu & Reddy (2008) menganalisis kemiskinan di tiga negara (Tanzania, Nicaragua dan Vietnam) menggunakan estimator kernel. Jung (2009) menganalisis masalah kemiskinan di Amerika Serikat bagian selatan menggunakan spasial kernel adaptif.

Menurut Budiantara, Ratnasari, Ratna, & Zain (2015), model-model regresi nonparametrik maupun semiparametrik yang dikembangkan oleh para peneliti selama ini, jika ditelusuri lebih mendalam, pada dasarnya terdapat asumsi yang sangat berat dan mendasar pada modelnya. Masing-masing prediktor dalam regresi nonparametrik multiprediktor dianggap memiliki pola yang sama sehingga para peneliti memaksakan penggunaan hanya satu bentuk estimator model untuk semua variabel prediktornya. Oleh karena itu, menggunakan hanya satu bentuk estimator saja dalam berbagai bentuk pola hubungan data yang berbeda-beda tentu akan mengakibatkan estimator yang dihasilkan kurang cocok dengan pola data. Akibatnya estimasi model regresi menjadi kurang baik dan menghasilkan error

yang besar. Oleh karena itu, untuk mengatasi masalah tersebut beberapa peneliti telah mengembangkan estimator kurva regresi campuran nonparametrik dimana masing-masing pola data dalam model regresi nonparametrik dihampiri dengan estimator kurva yang sesuai. Estimator kurva regresi campuran nonparametrik ini diharapkan menjadi estimator yang tepat yang mampu mengestimasi pola data dengan baik. Beberapa peneliti yang pernah mengembangkan estimator kurva regresi campuran nonparametrik diantaranya adalah Budiantara, dkk (2015) yang mengembangkan estimator kurva regresi campuran nonparametrik spline truncated dan kernel. Selain itu Sudiarsa, Budiantara, Suhartono & Purnami (2015) mengombinasikan estimator deret fourier dan spline truncated dalam kurva regresi nonparametrik multivariabel.

## 1.2 Perumusan Masalah

Pada penelitian sebelumnya, Budiantara, dkk (2015) telah melakukan pengembangan estimator kurva regresi campuran nonparametrik pada model regresi nonparametrik yang memiliki dua variabel prediktor. Hubungan variabel respon dengan masing-masing variabel prediktor memiliki bentuk pola hubungan yang berbeda. Bentuk pola hubungan variabel respon dengan variabel prediktor pertama adalah pola datanya memiliki perilaku berubah-ubah pada setiap sub-sub interval tertentu, sedangkan bentuk pola hubungan variabel respon dengan variabel prediktor kedua adalah pola datanya tidak diketahui atau tidak mengikuti pola tertentu.

Dalam prakteknya, variabel prediktor dalam suatu model regresi bisa saja lebih dari dua. Oleh karena itu, jika variabel prediktor lebih dari dua dan terdapat dua bentuk pola hubungan, maka perlu dilakukan kajian lebih lanjut mengenai estimator kurva regresi campuran nonparametrik yang telah dilakukan oleh Budiantara, dkk (2015) tersebut. Penelitian ini akan melakukan kajian mengenai model regresi nonparametrik aditif lebih dari dua prediktor, dimana bentuk pola hubungan salah satu prediktor dengan variabel respon adalah pola datanya berubah-ubah pada setiap sub-sub interval tertentu, sedangkan bentuk hubungan variabel prediktor yang lain dengan variabel respon adalah pola datanya tidak diketahui atau tidak mengikuti pola tertentu. Secara teoritis, pola data yang memiliki perilaku

berubah-ubah pada setiap sub-sub interval tertentu dapat dihampiri dengan kurva regresi spline, sedangkan pola data yang tidak diketahui atau tidak mengikuti pola tertentu dapat dihampiri dengan kurva regresi kernel.

Pokok permasalahan dalam model regresi nonparametrik multiprediktor tersebut adalah bagaimana bentuk estimator kurva regresi campuran nonparametriknya. Selanjutnya, bagaimana penerapan estimator kurva regresi campuran nonparametrik tersebut pada data riil, yaitu data kemiskinan di Provinsi Papua sehingga diperoleh model regresi campuran nonparametrik terbaik.

### **1.3 Tujuan Penelitian**

Untuk menjawab pokok permasalahan tersebut, maka disusunlah tujuan dari penelitian ini yaitu

1. Melakukan kajian mengenai estimator kurva regresi campuran nonparametrik spline dan kernel dalam model regresi campuran nonparametrik multiprediktor aditif.
2. Melakukan kajian mengenai sifat estimator kurva regresi campuran nonparametrik spline dan kernel.
3. Mengeksplorasi data kemiskinan di Provinsi Papua dan memodelkannya menggunakan regresi campuran nonparametrik spline dan kernel, selanjutnya membuat simulasi data untuk melakukan prediksi persentase penduduk miskin kabupaten/kota.

### **1.4 Manfaat Penelitian**

Dengan dilakukannya penelitian ini, diharapkan bisa diperoleh manfaat sebagai berikut.

1. Memberikan kontribusi dalam mengembangkan wawasan keilmuan bidang statistika khususnya pada regresi nonparametrik pendekatan spline dan kernel.
2. Memberikan model alternatif untuk menganalisis masalah kemiskinan yaitu dengan menggunakan model regresi campuran nonparametrik.

## 1.5 Batasan Penelitian

Dalam penelitian ini, ruang lingkup permasalahan dibatasi pada beberapa hal sebagai berikut.

1. Data yang digunakan adalah data makro seluruh kabupaten/kota di Provinsi Papua kondisi tahun 2013 yang bersumber dari Badan Pusat Statistik (BPS). Banyaknya kabupaten/kota di Provinsi Papua kondisi tahun 2013 adalah 29 kabupaten/kota. Konsep dan definisi data yang digunakan sama dengan konsep dan definisi yang digunakan oleh BPS.
2. Model yang digunakan adalah model regresi campuran nonparametrik spline linier truncated dan kernel Nadaraya-Watson, dimana pada model tersebut diasumsikan bahwa satu pola data mengikuti kurva spline, sedangkan beberapa pola data mengikuti kurva kernel.
3. Fungsi kernel yang digunakan adalah fungsi kernel Gaussian dan metode yang digunakan dalam menentukan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum adalah metode *Generalized Cross Validation* (GCV). Pemilihan banyaknya titik knot dibatasi sampai dengan tiga titik knot.

## BAB 2

### KAJIAN PUSTAKA

#### 2.1 Analisis Regresi

Analisis regresi digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor. Jika variabel respon adalah  $y_i$  dan variabel prediktor adalah  $x_i, i = 1, 2, \dots, n$ , maka pasangan data  $x_i, y_i$  akan memiliki model hubungan fungsional

$$y_i = f(x_i) + \varepsilon_i, \quad (2.1)$$

dimana  $f(x_i)$  adalah kurva regresi dan  $\varepsilon_i$  adalah error random yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian  $\sigma^2$  (Eubank, 1999). Analisis regresi memiliki dua tujuan utama. Pertama, memberikan cara mengeksplorasi hubungan antara variabel respon dan variabel prediktor dan yang kedua adalah membuat prediksi (Silverman, 1985).

#### 2.2 Regresi Parametrik, Nonparametrik dan Semiparametrik

Model regresi umumnya dibagi ke dalam tiga bentuk, yaitu model regresi parametrik, model regresi nonparametrik dan model regresi semiparametrik. Bentuk model regresi tersebut 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 (Islamiyati & Budiantara, 2007). 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 (Lestari & Budiantara, 2010). Dalam beberapa kasus, sebagian bentuk pola data diketahui, sedangkan untuk sebagian yang lain bentuk pola datanya tidak diketahui. Pada kasus ini, pendekatan model regresi yang disarankan adalah pendekatan model regresi semiparametrik (Wahba, 1990).

Menurut Budiantara (2009), dalam regresi parametrik terdapat asumsi yang sangat kaku dan kuat yaitu bentuk kurva regresi diketahui, misalnya linear, kuadratik, kubik, polinomial derajat- $p$ , eksponen, dan lain-lain. Untuk memodelkan

data menggunakan regresi parametrik linear, kuadrat, kubik atau yang lain, umumnya dimulai dengan membuat diagram pencar. Pendekatan regresi parametrik memiliki sifat yang sangat baik dari pandangan statistika inferensi, seperti sederhana, mudah interpretasinya, parsimoni, estimatornya tidak bias, tergolong estimator linear, efisien, konsisten, BLUE (*Best Linear Unbiased Estimator*), yang sangat jarang dimiliki oleh pendekatan regresi lain seperti regresi nonparametrik dan regresi semiparametrik. Dalam pandangan regresi nonparametrik, data mencari sendiri bentuk estimasi dari kurva regresinya tanpa harus dipengaruhi oleh faktor subyektifitas si peneliti. Ini berarti pendekatan model regresi nonparametrik sangatlah fleksibel dan sangat obyektif. Pendekatan model regresi parametrik yang sering digunakan oleh para peneliti adalah pendekatan model regresi linier sederhana dan model regresi linier berganda, sedangkan pada model regresi nonparametrik, pendekatan model regresi yang sering digunakan adalah pendekatan regresi spline dan regresi kernel.

Regresi nonparametrik umumnya tidak mensyaratkan uji asumsi klasik. Misalnya pada regresi nonparametrik ridge yang tidak menyarankan terjadinya multikolinieritas. Pelanggaran asumsi multikolinieritas biasanya terjadi pada regresi parametrik linier berganda dimana di antara variabel prediktor terdapat hubungan linier, sehingga pada estimasi parameter

$$\hat{\alpha} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \tilde{y} \quad (2.2)$$

menyebabkan determinan  $\mathbf{X}^T \mathbf{X}$  sama dengan 0 (matriks singular) atau determinan  $\mathbf{X}^T \mathbf{X}$  mendekati 0 (hampir singular). Adanya multikolinieritas menyebabkan terdapat akar karakteristik yang kecil dalam matriks  $\mathbf{X}^T \mathbf{X}$ , yang mengakibatkan varian  $\hat{\alpha}$  menjadi besar. Multikolinieritas yang tinggi mengakibatkan estimator  $\hat{\alpha}$  yang dihasilkan dari metode *Ordinary Least Square* (OLS) menjadi tidak stabil (peka terhadap perubahan kecil pada data yang kelihatannya tidak penting).

Pada regresi nonparametrik ridge, estimasi parameter (2.2) dimodifikasi menjadi

$$\hat{\alpha}_R = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \tilde{y},$$

dimana  $0 \leq \lambda \leq \infty$ ,  $\lambda$  adalah nilai konstan yang dipilih sebagai indeks dari kelas estimator. Adanya konstanta  $\lambda$  menyebabkan matriks  $\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}$  memiliki

determinan tidak sama dengan 0 (bukan matriks singular) atau tidak mendekati 0. Oleh karena itu, regresi nonparametrik ridge tidak mensyaratkan uji asumsi multikolinieritas, namun demikian estimator ridge bersifat bias, yaitu sebesar

$$Bias = \lambda(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \tilde{\alpha}.$$

Sifat bias pada regresi nonparametrik ridge ini memang sifat estimator yang baik yaitu tak bias, namun estimator ridge dapat mengatasi adanya multikolinieritas. Sifat tidak bias merupakan sifat yang dibutuhkan namun tidak terlalu penting. Hal ini disebabkan karena sifat tak bias tidak menunjukkan apapun mengenai penyebaran dari distribusi estimator. Suatu estimator yang tidak bias namun mempunyai varians yang besar seringkali menghasilkan estimasi yang jauh berbeda dari nilai parameter yang sebenarnya.

### 2.3 Regresi Parametrik Linier Berganda

Model regresi parametrik memiliki sifat yang sangat baik dari sisi statistika inferensia seperti sederhana, mudah interpretasinya, estimatornya tidak bias, efisien dan konsisten (Budiantara, 2009). Pada model regresi parametrik, apabila bentuk pola data cenderung mengikuti pola linier, maka pendekatan yang sesuai untuk bentuk pola data tersebut adalah pendekatan regresi linier. Begitu juga untuk bentuk pola data kuadratik menggunakan pendekatan regresi kuadratik dan bentuk pola data kubik menggunakan pendekatan regresi kubik (Härdle, 1994). Dalam banyak penelitian, model regresi linier berganda adalah model regresi parametrik yang paling sering digunakan. Pada regresi ini, variabel respon diprediksi berdasarkan asumsi bahwa variabel respon tersebut memiliki hubungan yang linier dengan variabel-variabel prediktor (Rencher & Schaalje, 2008).

Diberikan data berpasangan  $(t_{1i}, t_{2i}, \dots, t_{mi}, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana pasangan data tersebut memiliki hubungan yang diasumsikan mengikuti model regresi parametrik linier berganda. Pada model regresi tersebut, yang merupakan variabel respon adalah  $y_i$  dan variabel prediktor adalah  $t_{1i}, t_{2i}, \dots, t_{mi}$ . Bentuk fungsional model regresi tersebut adalah

$$y_i = \alpha_0 + \alpha_1 t_{1i} + \alpha_2 t_{2i} + \dots + \alpha_m t_{mi} + \varepsilon_i. \quad (2.3)$$

Parameter  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m$  merupakan parameter-parameter yang tidak diketahui, sedangkan  $\varepsilon_i$  adalah error random yang identik, independen dan berdistribusi normal dengan mean nol dan varian  $\sigma^2$ .

Persamaan (2.3) berlaku untuk  $i = 1$  sampai dengan  $i = n$ , sehingga kumpulan persamaan-persamaan (2.3) dari  $i = 1$  sampai dengan  $i = n$  dapat membentuk persamaan vektor

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \alpha_0 + \alpha_1 t_{11} + \alpha_2 t_{21} + \dots + \alpha_m t_{m1} \\ \alpha_0 + \alpha_1 t_{12} + \alpha_2 t_{22} + \dots + \alpha_m t_{m2} \\ \vdots \\ \alpha_0 + \alpha_1 t_{1n} + \alpha_2 t_{2n} + \dots + \alpha_m t_{mn} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Jika disederhanakan akan menjadi

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & t_{11} & t_{21} & \dots & t_{m1} \\ 1 & t_{12} & t_{22} & \dots & t_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{1n} & t_{2n} & \dots & t_{mn} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Dalam bentuk notasi matriks, dapat ditulis menjadi

$$\tilde{y} = \mathbf{T}\tilde{\alpha} + \tilde{\varepsilon}, \quad (2.4)$$

dimana

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{T} = \begin{bmatrix} 1 & t_{11} & t_{21} & \dots & t_{m1} \\ 1 & t_{12} & t_{22} & \dots & t_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{1n} & t_{2n} & \dots & t_{mn} \end{bmatrix}, \tilde{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}, \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Vektor  $\tilde{y}$  berukuran  $n \times 1$ , matriks  $\mathbf{T}$  berukuran  $n \times (m + 1)$ , vektor  $\tilde{\alpha}$  berukuran  $(m + 1) \times 1$  dan vektor  $\tilde{\varepsilon}$  berukuran  $n \times 1$ . Parameter  $\tilde{\alpha}$  pada persamaan (2.4) dapat diestimasi dengan menggunakan metode *Ordinary Least Square* (OLS) atau *Maximum Likelihood Estimation* (MLE). Estimator yang dihasilkan dari kedua persamaan tersebut sesuai dengan persamaan (2.2).

## 2.4 Regresi Nonparametrik Spline Truncated

Regresi spline merupakan suatu regresi polinomial dimana segmen-segmen polinomial yang berbeda digabungkan bersama pada titik-titik knot  $\xi_1, \xi_2, \dots, \xi_q$  dan kontinu sehingga bersifat fleksibel dibandingkan polinomial biasa. Kurva regresi spline mempunyai titik knot yang maksudnya adalah titik perpaduan

bersama dimana terjadi perubahan perilaku kurva. Kurva regresi spline dapat menyesuaikan diri secara efektif terhadap perubahan perilaku data, sehingga didapatkan hasil yang mendekati kebenaran (Eubank, 1988). Menurut Cox & O'Sullivan (1996), Kurva regresi spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu. Menurut Budiantara (2009), model regresi spline mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik.

Diberikan data berpasangan  $(u_i, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana  $y_i$  adalah variabel respon dan  $u_i$  adalah variabel prediktor. Pola hubungan  $y_i$  dan  $u_i$  dapat dinyatakan dalam model regresi

$$y_i = g(u_i) + \varepsilon_i . \quad (2.5)$$

Selanjutnya jika kurva regresi  $g(u_i)$  pada persamaan (2.5) dihampiri dengan kurva regresi spline truncated, maka

$$\begin{aligned} g(u_i) &= \beta_0 + \beta_1 u_i + \dots + \beta_p u_i^p + \lambda_1 (u_i - \xi_1)_+^p + \dots + \lambda_q (u_i - \xi_q)_+^p \\ &= \sum_{k=0}^p \beta_k u_i^k + \sum_{l=1}^q \lambda_l (u_i - \xi_l)_+^p , \end{aligned} \quad (2.6)$$

dimana

$$(u_i - \xi_l)_+^p = \begin{cases} (u_i - \xi_l)^p & , u_i \geq \xi_l \\ 0 & , u_i < \xi_l . \end{cases}$$

Kurva regresi  $g(u_i)$  merupakan kurva regresi nonparametrik spline truncated derajat  $p$  dengan banyaknya titik knot  $q$ . Derajat  $p$  merupakan derajat pada persamaan polinomial. Biasanya digunakan  $p = 1, 2$  atau  $3$ , karena untuk derajat kurva regresi di atasnya (lebih dari  $3$ ) bisa didekati dengan bentuk kurva polinomial tersebut. Kurva regresi polinomial derajat  $1$  biasa disebut dengan kurva regresi linier, kurva regresi polinomial derajat  $2$  biasa disebut dengan kurva regresi kuadratik, sedangkan kurva regresi polinomial derajat  $3$  biasa disebut dengan kurva regresi kubik. Titik-titik knot  $\xi_1, \xi_2, \dots, \xi_q$  adalah titik-titik knot yang menunjukkan pola perilaku dari kurva pada sub-sub interval yang berbeda, dimana  $\xi_1 < \xi_2 < \dots < \xi_q$ . Jika pendekatan regresi spline truncated memiliki derajat  $p = 1$ , maka akan

menjadi regresi spline linier truncated, dimana bentuk model regresi (2.6) akan menjadi

$$\begin{aligned} g(u_i) &= \beta_0 + \beta_1 u_i + \lambda_1 (u_i - \xi_1)_+ + \cdots + \lambda_q (u_i - \xi_q)_+ \\ &= \beta_0 + \beta_1 u_i + \sum_{l=1}^q \lambda_l (u_i - \xi_l)_+ \end{aligned}$$

Persamaan (2.6) berlaku untuk  $i = 1$  sampai dengan  $i = n$ , sehingga kumpulan persamaan-persamaan (2.6) dari  $i = 1$  sampai dengan  $i = n$  tersebut dapat digabungkan ke dalam bentuk persamaan vektor dan matriks

$$\begin{bmatrix} g(u_1) \\ g(u_2) \\ \vdots \\ g(u_n) \end{bmatrix} = \begin{bmatrix} 1 & u_1 & \cdots & u_1^p & (u_1 - \xi_1)_+^p & \cdots & (u_1 - \xi_q)_+^p \\ 1 & u_2 & \cdots & u_2^p & (u_2 - \xi_1)_+^p & \cdots & (u_2 - \xi_q)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_n & \cdots & u_n^p & (u_n - \xi_1)_+^p & \cdots & (u_n - \xi_q)_+^p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_q \end{bmatrix}.$$

Dalam notasi matriks dapat ditulis menjadi

$$\tilde{g}(u) = \mathbf{G}(\tilde{\xi})\tilde{\theta}, \quad (2.7)$$

dimana

$$\tilde{g}(u) = \begin{bmatrix} g(u_1) \\ g(u_2) \\ \vdots \\ g(u_n) \end{bmatrix}, \quad \tilde{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_q \end{bmatrix},$$

$$\mathbf{G}(\tilde{\xi}) = \begin{bmatrix} 1 & u_1 & \cdots & u_1^p & (u_1 - \xi_1)_+^p & \cdots & (u_1 - \xi_q)_+^p \\ 1 & u_2 & \cdots & u_2^p & (u_2 - \xi_1)_+^p & \cdots & (u_2 - \xi_q)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_n & \cdots & u_n^p & (u_n - \xi_1)_+^p & \cdots & (u_n - \xi_q)_+^p \end{bmatrix}.$$

Vektor  $\tilde{g}(u)$  berukuran  $n \times 1$ , matriks  $\mathbf{G}(\tilde{\xi})$  berukuran  $n \times (p + q + 1)$ , sedangkan vektor  $\tilde{\theta}$  berukuran  $(p + q + 1) \times 1$ .

Sebelum estimasi terhadap kurva regresi  $\tilde{g}(u)$  pada persamaan (2.7), terlebih dahulu dilakukan estimasi terhadap parameter  $\tilde{\theta}$ . Estimasi parameter  $\tilde{\theta}$  dilakukan dengan menggunakan metode MLE. Fungsi likelihood yang diberikan adalah

$$\begin{aligned} L(\tilde{\theta}, \sigma^2 | \tilde{\xi}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \varepsilon_i^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|\tilde{\varepsilon}\|^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2\right). \end{aligned}$$

Estimator MLE yang dihasilkan adalah

$$\hat{\theta}(\tilde{\xi}) = \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T \tilde{y}.$$

Dengan demikian, estimasi dari kurva regresi  $\tilde{g}(u)$  adalah

$$\begin{aligned} \hat{g}_{\tilde{\xi}}(u) &= \mathbf{G}(\tilde{\xi}) \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T \tilde{y} \\ &= \mathbf{S}(\tilde{\xi}) \tilde{y}, \end{aligned}$$

dimana:

$$\mathbf{S}(\tilde{\xi}) = \mathbf{G}(\tilde{\xi}) \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T.$$

Vektor  $\hat{g}_{\tilde{\xi}}(u)$  dan  $\hat{\theta}(\tilde{\xi})$  memiliki ukuran yang sama dengan vektor parameternya.

Sedangkan matriks  $\mathbf{S}(\tilde{\xi})$  berukuran  $n \times (p + q + 1)$ .

Titik-titik knot merupakan pengontrol keseimbangan antara kemulusan kurva dan kesesuaian kurva terhadap data. Pemilihan banyaknya titik knot, lokasi titik-titik knot optimum dapat dilakukan salah satunya dengan metode *Generalized Cross Validation* atau GCV (Eubank, 1988).

$$GCV(\tilde{\xi}) = \frac{MSE(\tilde{\xi})}{\left(n^{-1} \text{tr}(\mathbf{I} - \mathbf{S}(\tilde{\xi}))\right)^2}, \quad (2.8)$$

dimana  $MSE(\tilde{\xi})$  pada persamaan (2.8) adalah

$$MSE(\tilde{\xi}) = n^{-1} \sum_{i=1}^n (y_i - \hat{g}(u_i))^2.$$

Pemilihan banyaknya titik knot, lokasi titik-titik knot optimum dilakukan dengan melihat nilai GCV yang minimum.

## 2.5 Regresi Nonparametrik Kernel

Diberikan pasangan pengamatan independen  $(v_i, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana  $y_i$  adalah variabel respon, sedangkan  $v_i$  adalah variabel prediktor. Hubungan  $y_i$  dan  $v_i$  dapat dimodelkan secara fungsional dalam bentuk

$$y_i = h(v_i) + \varepsilon_i, \quad (2.9)$$

dimana kurva regresi  $h(v_i)$  merupakan kurva yang tidak diketahui bentuknya. Menurut Hardle (1994), kurva  $h(v_i)$  pada model nonparametrik tersebut dapat diestimasi dengan pendekatan kernel yang didasarkan pada fungsi densitas kernel.

Kurva  $h(v_i)$  pada model (2.9) dapat diestimasi menggunakan estimator kernel Nadaraya-Watson. Estimator kernel Nadaraya-Watson merupakan kasus khusus dari kurva regresi polinomial lokal yaitu kurva regresi polinomial lokal yang memiliki orde sama dengan 0 atau disebut juga dengan kurva regresi konstan lokal. Ketika kurva regresi polinomial lokal memiliki orde sama dengan 1, maka kurva regresi polinomial lokal disebut juga dengan kurva regresi linier lokal.

Kurva regresi polinomial lokal mengadopsi dari perluasan deret Taylor di sekitar  $v$ . Jika  $h(v_i)$  pada persamaan (2.9) dapat dihampiri dengan kurva regresi polinomial lokal

$$\begin{aligned} h(v_i) &= \beta_0 + \beta_1(v_i - v) + \beta_2(v_i - v)^2 + \dots + \beta_p(v_i - v)^p \\ &= \sum_{k=0}^p \beta_k(v_i - v)^k, \end{aligned}$$

dimana

$$\beta_k = \frac{h^k(v)}{k!}, \quad k = 0, 1, 2, \dots, p.$$

Dalam bentuk matriks dapat ditulis

$$\tilde{y} = \mathbf{V}\tilde{\beta} + \tilde{\varepsilon},$$

dimana

$$\mathbf{V} = \begin{bmatrix} 1 & (v_1 - v) & (v_1 - v)^2 & \dots & (v_1 - v)^p \\ 1 & (v_2 - v) & (v_2 - v)^2 & \dots & (v_2 - v)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (v_n - v) & (v_n - v)^2 & \dots & (v_n - v)^p \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Parameter  $\tilde{\beta}$  bergantung pada titik  $v$  yang disebut sebagai titik lokal. Parameter  $\tilde{\beta}$  dapat diestimasi menggunakan *weighted least square* (WLS) dengan meminimumkan fungsi

$$L = \sum_{i=1}^n \left( y_i - \sum_{k=0}^p \beta_k (v_i - v)^k \right)^2 K\left(\frac{v_i - v}{\phi}\right), \quad (2.10)$$

dimana  $\phi$  merupakan bandwidth yang mengontrol ukuran persekitaran titik lokal  $v_0$ . Permasalahan *weighted least square* pada (2.10) dalam bentuk matriks dapat ditulis menjadi

$$L = (\tilde{y} - \mathbf{V}\tilde{\beta})^T \mathbf{W}(\tilde{y} - \mathbf{V}\tilde{\beta}),$$

dimana  $\mathbf{W}$  merupakan matriks diagonal yang berukuran  $n \times n$ ,

$$\mathbf{W} = \text{diag}\left\{K\left(\frac{v_i - v}{\phi}\right)\right\}.$$

Hasil estimasi dari parameter  $\tilde{\beta}$  adalah

$$\hat{\tilde{\beta}} = (\mathbf{V}^T \mathbf{W} \mathbf{V})^{-1} \mathbf{V}^T \mathbf{W} \tilde{y},$$

sehingga estimasi model regresi polinomial lokal adalah

$$\begin{aligned} \hat{y} &= \mathbf{V}(\mathbf{V}^T \mathbf{W} \mathbf{V})^{-1} \mathbf{V}^T \mathbf{W} \tilde{y} \\ &= \mathbf{H} \tilde{y}, \end{aligned}$$

dimana

$$\mathbf{H} = \mathbf{V}(\mathbf{V}^T \mathbf{W} \mathbf{V})^{-1} \mathbf{V}^T \mathbf{W}.$$

Terkait dengan model regresi polinomial lokal, model regresi kernel Nadaraya-Watson merupakan model regresi polinomial lokal yang hanya memuat konstan lokal. Jadi jika  $h(v_i)$  hanya memuat konstan lokal, maka dengan meminimumkan fungsi

$$L = \sum_{i=1}^n (y_i - \beta_0)^2 K\left(\frac{v_i - v}{\phi}\right),$$

akan dihasilkan

$$\beta_0 = \sum_{i=1}^n \frac{K\left(\frac{v_i - v}{\phi}\right)}{\sum_{i=1}^n K\left(\frac{v_i - v}{\phi}\right)} y_i,$$

sehingga

$$\hat{h}_\phi(v_i) = n^{-1} \sum_{i=1}^n \frac{K_\phi(v_i - v)}{n^{-1} \sum_{i=1}^n K_\phi(v_i - v)} y_i$$

$$= n^{-1} \sum_{i=1}^n W_{\phi i}(\nu) y_i \quad (2.11)$$

Fungsi  $W_{\phi j i}(\nu_j)$  merupakan fungsi pembobot

$$W_{\phi i}(\nu) = \frac{K_\phi(\nu_i - \nu)}{n^{-1} \sum_{i=1}^n K_\phi(\nu_i - \nu)}$$

dimana  $K_{\phi j}(\nu_j - \nu_{ji})$  adalah fungsi kernel

$$K_\phi(\nu_i - \nu) = \frac{1}{\phi} K\left(\frac{\nu_i - \nu}{\phi}\right).$$

Fungsi kernel  $K$  adalah fungsi yang bernilai riil, kontinu, terbatas dan simetris dengan integralnya sama dengan satu atau  $\int K(z) dz = 1$ . Fungsi kernel  $K$  dapat berupa kernel uniform, kernel segitiga, kernel epanechnikov, kernel kuadrat, kernel triweight, kernel kosinus atau kernel Gaussian (Härdle, Müller, Sperlich, & Werwatz, 2004). Kernel Gaussian cukup sering digunakan dalam banyak penelitian. Fungsi kernel Gaussian lebih *smooth* dibandingkan dengan fungsi kernel yang lain. Bentuk fungsi kernel Gaussian adalah

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right), \quad (2.12)$$

dimana  $-\infty < z < \infty$ .

Jika bentuk penjumlahan pada persamaan (2.11) dijabarkan dengan lebih lengkap, maka

$$\hat{h}_\phi(\nu_i) = n^{-1} W_{\phi 1}(\nu) y_1 + n^{-1} W_{\phi 2}(\nu) y_2 + \cdots + n^{-1} W_{\phi n}(\nu) y_n$$

Karena berlaku untuk setiap  $\nu = \nu_1$  sampai dengan  $\nu = \nu_n$ , maka jika persamaan untuk  $\nu_1$  sampai  $\nu_n$  digabungkan ke dalam bentuk matriks akan menjadi

$$\begin{bmatrix} \hat{h}_\phi(\nu_1) \\ \hat{h}_\phi(\nu_2) \\ \vdots \\ \hat{h}_\phi(\nu_n) \end{bmatrix} = \begin{bmatrix} n^{-1} W_{\phi 1}(\nu_1) & n^{-1} W_{\phi 2}(\nu_1) & \cdots & n^{-1} W_{\phi n}(\nu_1) \\ n^{-1} W_{\phi 1}(\nu_2) & n^{-1} W_{\phi 2}(\nu_2) & \cdots & n^{-1} W_{\phi n}(\nu_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\phi 1}(\nu_n) & n^{-1} W_{\phi 2}(\nu_n) & \cdots & n^{-1} W_{\phi n}(\nu_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Jika dinotasikan ke dalam bentuk matriks akan menjadi

$$\hat{h}_\phi(\nu) = \mathbf{V}(\boldsymbol{\phi}) \tilde{\mathbf{y}},$$

dimana

$$\hat{h}_\phi(v) \begin{bmatrix} \hat{h}_\phi(v_1) \\ \hat{h}_\phi(v_2) \\ \vdots \\ \hat{h}_\phi(v_n) \end{bmatrix},$$

$$\mathbf{V}(\phi) = \begin{bmatrix} n^{-1}W_{\phi 1}(v_1) & n^{-1}W_{\phi 2}(v_1) & \dots & n^{-1}W_{\phi n}(v_1) \\ n^{-1}W_{\phi 1}(v_2) & n^{-1}W_{\phi 2}(v_2) & \dots & n^{-1}W_{\phi n}(v_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_{\phi 1}(v_n) & n^{-1}W_{\phi 2}(v_n) & \dots & n^{-1}W_{\phi n}(v_n) \end{bmatrix}.$$

Vektor  $\hat{h}_\phi(v)$  berukuran  $n \times 1$ , sedangkan matriks  $\mathbf{V}(\phi)$  berukuran  $n \times n$ .

Estimator kernel sangat sensitif pada pemilihan bandwidth  $\phi$  yang fungsinya mengontrol kemulusan kurva estimasi. Jika bandwidth  $\phi$  diperkecil maka bias akan turun, namun varian akan membesar, sebaliknya jika bandwidth  $\phi$  diperbesar maka varian yang akan turun, namun bias akan membesar. Oleh karena itu, perlu adanya keseimbangan antara bias dan varian melalui pemilihan bandwidth  $\phi$  yang optimum (tidak terlalu kecil dan tidak terlalu besar). Penggunaan bandwidth  $\phi$  yang optimum akan menghasilkan *mean squared error* (MSE) terkecil. Hubungan MSE, bias dan varian adalah

$$\begin{aligned} \text{MSE}[\hat{h}_\phi(v_i)] &= E[\hat{h}_\phi(v_i) - h_\phi(v_i)]^2 \\ &= E[\hat{h}_\phi(v_i) - E[\hat{h}_\phi(v_i)]]^2 + (E[\hat{h}_\phi(v_i)] - h_\phi(v_i))^2 \\ &= \text{Var}[\hat{h}_\phi(v_i)] + (\text{Bias}[\hat{h}_\phi(v_i)])^2. \end{aligned}$$

Salah satu metode yang digunakan untuk mendapatkan bandwidth optimum adalah metode GCV.

$$\text{GCV}(\phi) = \frac{\text{MSE}(\phi)}{\left(n^{-1}\text{tr}(\mathbf{I} - \mathbf{V}(\phi))\right)^2}, \quad (2.13)$$

dimana  $\text{MSE}(\phi)$  pada persamaan (2.13) adalah

$$\text{MSE}(\phi) = n^{-1} \sum_{i=1}^n (y_i - \hat{h}_\phi(v_i))^2.$$

Nilai bandwidth yang optimum berkaitan dengan nilai GCV yang minimum (Eubank, 1999).

## 2.6 Regresi Campuran 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) adalah peneliti yang pernah mengembangkan model regresi campuran nonparametrik spline dan kernel. Model yang dikembangkannya menggunakan satu variabel respon dan dua variabel prediktor. Pola hubungan variabel respon dengan variabel prediktor pertama diasumsikan mengikuti kurva regresi spline truncated, sedangkan pola hubungan variabel respon dan variabel prediktor kedua diasumsikan mengikuti kurva kernel Nadaraya-Watson.

Diberikan data berpasangan  $(u_i, v_i, y_i)$ ,  $i = 1, 2, \dots, n$ , dimana hubungan antara variabel prediktor  $u_i, v_i$  dengan variabel respon  $y_i$  diasumsikan mengikuti model regresi nonparametrik

$$y_i = f(u_i, v_i) + \varepsilon_i . \quad (2.14)$$

Bentuk kurva  $f(u_i, v_i)$  pada persamaan (2.14) diasumsikan tidak diketahui dan hanya diasumsikan *smooth* dalam arti kontinu dan differensiabel. Selanjutnya kurva regresi  $f(u_i, v_i)$  diasumsikan bersifat aditif sehingga dapat ditulis dalam bentuk

$$f(u_i, v_i) = g(u_i) + h(v_i) . \quad (2.15)$$

Pada persamaan (2.15), kurva regresi  $g(u_i)$  diasumsikan mengikuti kurva regresi spline truncated, sedangkan kurva regresi  $h(v_i)$  diasumsikan mengikuti kurva regresi kernel. Estimator kurva regresi campuran nonparametrik spline dan kernel tersebut adalah

$$\hat{f}_{\phi, \xi}(u, v) = \hat{g}_{\phi, \xi}(u, v) + \hat{h}_{\phi}(v) . \quad (2.16)$$

Parameter  $\phi$  pada persamaan (2.16) adalah parameter bandwidth dan  $\tilde{\xi} = (\xi_1, \xi_2, \dots, \xi_q)^T$  adalah titik-titik knot.

Menurut Budiantara, dkk (2015), estimator kurva regresi campuran nonparametrik spline truncated dan kernel pada persamaan (2.16) dapat diperoleh dengan menggunakan metode MLE. Fungsi likelihood yang diberikan adalah

$$L(\tilde{\theta}, \sigma^2 | \phi, \tilde{\xi}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \varepsilon_i^2\right)$$

$$\begin{aligned}
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|\tilde{\varepsilon}\|^2\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|(\mathbf{I} - \mathbf{V}(\boldsymbol{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2\right),
\end{aligned}$$

sehingga estimator MLE yang dihasilkan adalah

$$\hat{f}_{\phi, \tilde{\xi}}(u, v) = \mathbf{Z}(\tilde{\xi}, \boldsymbol{\phi})\tilde{y}$$

$$\hat{g}_{\phi, \tilde{\xi}}(u, v) = \mathbf{S}(\tilde{\xi}, \boldsymbol{\phi})\tilde{y}$$

$$\hat{h}_{\phi}(v) = \mathbf{V}(\boldsymbol{\phi})\tilde{y},$$

dimana

$$\mathbf{Z}(\tilde{\xi}, \boldsymbol{\phi}) = \mathbf{S}(\tilde{\xi}, \boldsymbol{\phi}) + \mathbf{V}(\boldsymbol{\phi})$$

$$\mathbf{S}(\tilde{\xi}, \boldsymbol{\phi}) = \mathbf{G}(\tilde{\xi})\hat{\theta}(\tilde{\xi}, \boldsymbol{\phi})$$

$$\hat{\theta}(\tilde{\xi}, \boldsymbol{\phi}) = \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\boldsymbol{\phi}))\tilde{y}.$$

Matriks  $\mathbf{Z}(\tilde{\xi}, \boldsymbol{\phi})$  dan matriks  $\mathbf{S}(\tilde{\xi}, \boldsymbol{\phi})$  adalah matrik berukuran  $n \times n$ .

Estimator campuran nonparametrik spline dan kernel  $\hat{f}_{\phi, \tilde{\xi}}(u, v)$  sangat tergantung pada lokasi knot, banyak titik-titik knot  $\tilde{\xi}$  dan parameter bandwidth  $\boldsymbol{\phi}$ . Untuk memperoleh estimator kurva regresi campuran nonparametrik spline dan kernel yang terbaik, perlu dilakukan pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum . Salah satu metode yang bisa digunakan untuk pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum adalah dengan metode GCV.

$$GCV(\tilde{\xi}, \boldsymbol{\phi}) = \frac{MSE(\tilde{\xi}, \boldsymbol{\phi})}{\left(n^{-1} \text{tr}(\mathbf{I} - \mathbf{Z}(\tilde{\xi}, \boldsymbol{\phi}))\right)^2}, \quad (2.17)$$

dimana  $MSE(\tilde{\xi}, \boldsymbol{\phi})$  pada persamaan (2.17) dihitung dengan

$$MSE(\tilde{\xi}, \boldsymbol{\phi}) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}_{\phi, \tilde{\xi}}(u_i, v_i))^2.$$

Banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum diperoleh dari nilai GCV yang minimum. Pemilihan lokasi titik-titik knot dan bandwidth dilakukan secara simultan.

## 2.7 Optimasi nlminb

Fungsi nlminb pada *software R* merupakan fungsi optimasi *unconstrained* dan *box-constrained* menggunakan PORT routines yang ditulis oleh Gay (1990) di Bell Labs, selanjutnya diimporkan ke *software R* oleh Douglas Bates. Fungsi nlminb menggunakan metode optimasi Newton (Nash dan Varadhan, 2011). Beberapa optimisasi PORT routines memiliki struktur dan *user interface* yang umum. Semua kontrol, toleransi dan semua penyimpanan awal yang digunakan oleh routines terkandung dalam dua array: IV untuk solusi yang bernilai *integer*, V untuk solusi yang bernilai *floating-point* (REAL dalam versi presisi tunggal, DOUBLE PRECISION dalam versi presisi ganda).

Diberikan suatu fungsi  $f$  dari sebanyak  $p$  variabel, optimasi routines mencoba untuk menemukan  $p$ -vektor  $\tilde{\phi}$  yang meminimumkan fungsi  $f(\tilde{\phi})$ .

Berbagai macam batasan bisa dilakukan pada  $\tilde{\phi}$ , yaitu

- a. tidak ada batasan,
- b. batasan sederhana  $\underline{b}^{\tilde{\phi}} \leq \tilde{\phi} \leq \bar{b}^{\tilde{\phi}}$ , dimana  $\underline{b}^{\tilde{\phi}}$  dan  $\bar{b}^{\tilde{\phi}}$  adalah vektor, dan
- c. batasan linier umum  $\underline{b}^c \leq \mathbf{C}\tilde{\phi} \leq \bar{b}^c$ , dimana  $\underline{b}^c$  dan  $\bar{b}^c$  merupakan vektor dan  $\mathbf{C}$  adalah matriks.

Gradien  $f$  pada  $\tilde{\phi}$  (vektor turunan parsial pertama dari  $f$ ) dinotasikan oleh  $\nabla f(\tilde{\phi})$ , dan Hessian dari  $f$  pada  $\tilde{\phi}$  (matriks dari turunan parsial kedua dari  $f$ ) dinotasikan oleh  $\nabla^2 f(\tilde{\phi})$ .

Optimasi routines memiliki dua tingkat, yaitu *forward-communication routine* dan *reverse-communication iteration driver*. *Forward-communication routine* mempelajari tentang  $f(\tilde{\phi})$  dengan cara konvensional. Jika diberikan subroutines, maka *forward-communication routine* dapat melakukan panggilan untuk menghitung  $f(\tilde{\phi})$ . Di sisi lain, *reverse-communication driver* kembali melakukan pemanggilan ketika *reverse-communication driver* perlu mengetahui  $f(\tilde{\phi})$  pada  $\tilde{\phi}$  yang baru. Pemanggilan routines harus menghitung informasi yang diperlukan (misalnya  $f(\tilde{\phi})$  itu sendiri, atau untuk beberapa routines regresi, vektor residual) dan memanggil *reverse-communication driver* kembali. Biasanya lebih mudah untuk menggunakan optimasi *forward-communication routine*, tapi kadang-

kadang lebih mudah untuk memanggil optimasi *reverse-communication driver*, misalnya ketika penulisan sebuah *subroutine* yang menghitung  $f(\tilde{\phi})$  sulit.

Kedua versi *forward-communication* dan *reverse-communication* dari optimasi routines menerima IV dan V sebagai parameter. Parameter untuk *reverse-communication drivers* juga termasuk  $\tilde{\phi}$ , salah satu dari nilai  $f(\tilde{\phi})$  dan mungkin  $\nabla f(\tilde{\phi})$  atau informasi yang cukup untuk menghitung nilai-nilai ini (untuk routines regresi). *Driver* kembali dengan  $\text{IV}(1) = 1$  ketika menginginkan memiliki  $f$  yang dievaluasi pada  $\tilde{\phi}$  saat ini dan dengan  $\text{IV}(1) = 2$  ketika menginginkan  $\nabla f$  yang dievaluasi. Beberapa drivers memiliki kemungkinan lain untuk kembali, seperti pada  $\text{IV}(1) = -1$  atau  $-2$ .

## 2.8 Uji Normalitas Kolmogorov-Smirnov

Normalitas residual diperiksa menggunakan uji Kolmogorov-Smirnov dengan hipotesis

$$H_0 : F(\varepsilon) = F_0(\varepsilon) \text{ (Residual berdistribusi Normal)}$$

$$H_1 : F(\varepsilon) \neq F_0(\varepsilon) \text{ (Residual tidak berdistribusi Normal)}.$$

Statistik uji yang digunakan adalah

$$D = \max |F_0(\varepsilon) - S_n(\varepsilon)|,$$

dimana  $F_0(\varepsilon)$  adalah fungsi distribusi kumulatif normal dan  $S_n(\varepsilon)$  merupakan fungsi peluang kumulatif data sampel. Sedangkan  $D$  merupakan nilai maksimum semua  $\varepsilon$  dari nilai mutlak selisih antara nilai  $F_0(\varepsilon)$  dengan  $S_n(\varepsilon)$ . Tolak  $H_0$  apabila  $D > q_{(1-\alpha)}$ , dimana nilai  $q_{(1-\alpha)}$  didasarkan pada tabel Kolmogorov-Smirnov atau error berdistribusi normal ketika  $p\text{-value} > \alpha$ .

## 2.9 Kriteria Kebaikan Model

Nilai koefisien determinasi  $R^2$  (*coefficient of determination /R-squared*), merupakan salah satu kriteria kebaikan model. Nilai  $R^2$  yang tinggi adalah kriteria nilai model yang baik untuk meramalkan data. Nilai  $R^2$  didefinisikan oleh

$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}},$$

dengan SSR adalah *Sum of Square Regression* dan SST adalah *Sum of Square Total*, dimana

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 , \quad SST = \sum_{i=1}^n (y_i - \bar{y})^2 .$$

## 2.10 Tinjauan Kemiskinan

Pada bagian ini diuraikan mengenai konsep dan definisi kemiskinan, pengukuran kemiskinan serta faktor-faktor yang mempengaruhi kemiskinan.

### 2.10.1 Konsep dan Pengukuran Kemiskinan

Dalam dimensi ekonomi, kemiskinan diartikan sebagai kekurangan sumber daya yang dapat digunakan untuk meningkatkan kesejahteraan sekelompok orang, baik secara finansial maupun semua jenis kekayaan yang dapat meningkatkan kesejahteraan masyarakat. Dikategorikan miskin bilamana seseorang atau keluarga tidak dapat memenuhi kebutuhan pokok minimumnya sandang, pangan, papan, kesehatan, dan pendidikan. Dimensi ekonomi dapat diukur dengan nilai rupiah meskipun harganya selalu berubah-ubah setiap tahunnya tergantung pada tingkat inflasi (Ellies, 1994).

Penetapan pengukuran dan kriteria kemiskinan secara nasional sangat sulit. Masih diperlukan kajian yang dapat mengakomodasikan permasalahan kemiskinan yang kompleks baik dari segi ekonomi, budaya, sosial, psikologi dan geografi yang sangat bervariasi di Indonesia. Hampir semua pendekatan dalam mengkaji kemiskinan masih berporos pada paradigma modernisasi (*modernization paradigm*) yang dimotori oleh Bank Dunia. Paradigma ini bersandar pada teori-teori pertumbuhan ekonomi neo klasik (*orthodox neoclassical economics*) dan model yang berpusat pada produksi (*production centered model*). Sejak pendapatan nasional (GNP) mulai dijadikan indikator pembangunan tahun 1950-an, para ilmuwan sosial selalu merujuk pada pendekatan tersebut manakala berbicara masalah kemiskinan satu negara. Pengukuran kemiskinan kemudian sangat dipengaruhi oleh perspektif *income poverty* yang menggunakan pendapatan sebagai satu-satunya indikator garis kemiskinan (Suharto 2009).

Untuk mengukur kemiskinan, BPS menggunakan konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*). Dengan pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Dengan kata lain, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan makanan maupun non makanan yang bersifat mendasar. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah Garis Kemiskinan. Garis Kemiskinan (GK) terdiri dari dua komponen yaitu Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). BPS (2014) mendefinisikan garis kemiskinan sebagai penjumlahan antara garis kemiskinan makanan dan garis kemiskinan non makanan. Garis Kemiskinan Makanan adalah nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2.100 kilokalori perkapita per hari. 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-lain). Sedangkan Garis Kemiskinan Non Makanan adalah 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.

### **2.10.2 Faktor-faktor yang Mempengaruhi Kemiskinan**

Sejumlah variabel dapat dipakai untuk melacak persoalan kemiskinan. Dari variabel-variabel tersebut dapat dihasilkan serangkaian strategi dan kebijakan penanggulangan kemiskinan yang tepat sasaran dan berkesinambungan. Dari dimensi pendidikan misalnya, pendidikan yang rendah dipandang sebagai penyebab kemiskinan. Dari dimensi kesehatan, rendahnya mutu kesehatan masyarakat menyebabkan terjadinya kemiskinan (Sahdan, 2005).

Penelitian yang dilakukan oleh Tilak (2005) menyatakan bahwa pendidikan berhubungan negatif terhadap kemiskinan dimana ketika tingkat pendidikan meningkat maka kemiskinan akan menurun. Pendidikan menjadi penting karena akan meningkatkan kapabilitas masyarakat. Apabila tingkat pendidikan masyarakat rendah, kapabilitas juga rendah. Hal ini akan berpengaruh

terhadap rendahnya kemampuan untuk menangkap peluang, sehingga akan berakibat pada kemiskinan. Pada banyak kasus, kondisi kemiskinan yang dialami oleh seseorang telah menyebabkan rendahnya pencapaian pendidikan atau terjadinya lingkaran setan (*vicious cycle*). Kenyataan inilah yang dialami mayoritas rakyat Indonesia (Naja, 2006).

Seseorang yang tidak bekerja sudah tentu tidak memperoleh pendapatan dan selanjutnya tidak akan mampu memenuhi kebutuhan hidupnya. Oleh karena itulah tingkat pengangguran menjadi salah satu faktor yang mempengaruhi tingkat kemiskinan. Penelitian Agenor (2004) melihat bahwa hubungan kemiskinan dan pengangguran berlawanan arah atau hubungannya negatif, di mana ketika pengangguran tinggi terjadi pada saat kemiskinan rendah. Hal ini terjadi disebabkan lebih besarnya ukuran sektor informal di negara-negara berkembang, sehingga tingkat pengangguran terselubung tidak termasuk dalam kategori pengangguran terbuka. Studi ini diaplikasikan untuk negara berkembang dengan menggunakan regresi antar negara. Sebaliknya hasil studi Formby, Hoover, & Hoseong (2000) menyimpulkan adanya hubungan positif dan signifikan antara tingkat pengangguran terbuka dan tingkat kemiskinan agregat di Amerika Serikat.

Menurut Fan (2003) untuk mengatasi kemiskinan, paling tidak ada dua faktor penting yang perlu diperhatikan yaitu pertumbuhan ekonomi dan distribusi pendapatan. Selain itu dikemukakan bahwa pengeluaran publik merupakan salah satu faktor yang dapat mempercepat pertumbuhan ekonomi yang akhirnya akan meningkatkan pendapatan dan mengurangi jumlah penduduk yang hidup di bawah garis kemiskinan. Oleh karena itu diperlukan upaya untuk melakukan pengurangan kemiskinan yaitu dengan adanya program dan kebijakan dari pemerintah daerah yang diharapkan mampu memberikan kesempatan luas bagi pelaku ekonomi di daerah terutama di kawasan yang sangat memerlukan perhatian pemerintah berupa kebijakan dan kemudahan dalam melakukan kegiatan ekonomi, sehingga dampak yang ditimbulkan dapat mendorong dan meningkatkan kemampuan ekonomi masyarakat miskin. Artinya bahwa strategi pembangunan ke depan harus diupayakan untuk menciptakan dan menggali potensi-potensi ekonomi agar kegiatan ekonomi yang dilakukan akan memacu peningkatan pertumbuhan

ekonomi serta mampu menurunkan perbedaan dan ketimpangan distribusi pendapatan masyarakat.

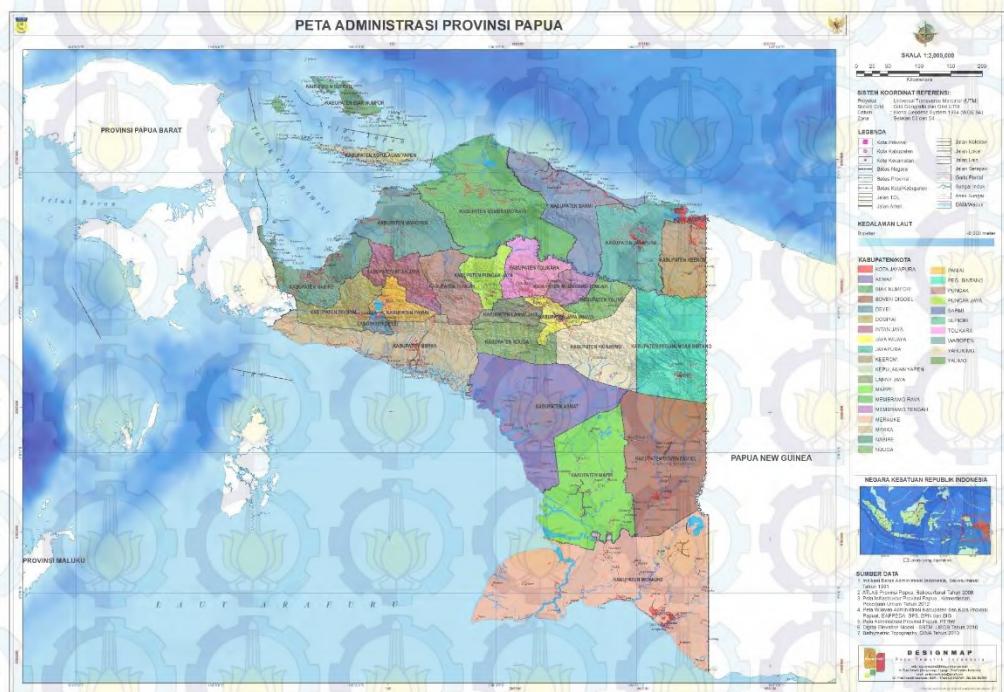
Menurut Boediono (1981) pertumbuhan ekonomi adalah proses kenaikan output per kapita dalam jangka panjang. Dalam definisi yang lebih ketat definisi tersebut ditekankan pada sumber dari dalam negeri. Sama halnya dengan daerah, pertumbuhan ekonomi akan dapat diwujudkan apabila pertumbuhan PDRB riil melebihi laju pertumbuhan penduduk. Untuk kepentingan analisis ekonomi regional biasanya digunakan pertumbuhan Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan. Jika laju pertumbuhan penduduk tetap, maka laju pertumbuhan PDRB riil akan meningkatkan pendapatan per kapita yang disertai pemerataan distribusinya sehingga dapat mengurangi kemiskinan. Beberapa studi empiris yang meneliti pengaruh pertumbuhan pendapatan per kapita terhadap tingkat kemiskinan antara lain dilakukan oleh Irawan (2005) dan Agrawal (2008).

# BAB 3

## METODE PENELITIAN

### 3.1 Sumber Data

Penelitian ini menggunakan data sekunder tahun 2013 yang diperoleh dari publikasi yang diterbitkan oleh Badan Pusat Statistik (BPS). Unit observasi yang digunakan adalah seluruh kabupaten/kota yang ada di Provinsi Papua, yaitu sebanyak 29 kabupaten/kota. Data kabupaten/kota yang digunakan disajikan pada Lampiran 1. Letak kabupaten/kota tersebut dapat dilihat dari peta Provinsi Papua pada Gambar 3.1.



Gambar 3.1 Peta Provinsi Papua

### **3.2 Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon dan lima variabel prediktor. Satu variabel respon tersebut adalah salah satu indikator kemiskinan, yaitu persentase penduduk miskin, sedangkan lima variabel prediktornya adalah variabel-variabel yang diduga mempengaruhi kemiskinan. Variabel yang diasumsikan mempengaruhi kemiskinan adalah rata-rata lama

sekolah, tingkat pengangguran terbuka (TPT), gini ratio, laju pertumbuhan ekonomi dan PDRB perkapita. Semua variabel tersebut akan diaplikasikan ke dalam model regresi campuran nonparametrik spline dan kernel.

Berikut ini adalah konsep dan definisi variabel operasional yang digunakan.

1. Persentase penduduk miskin adalah persentase penduduk yang berada di bawah garis kemiskinan. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah Garis Kemiskinan. Garis Kemiskinan (GK) terdiri dari dua komponen yaitu Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM).
2. Rata-rata lama sekolah adalah rata-rata jumlah tahun yang digunakan oleh penduduk usia 15 tahun ke atas dalam menjalani pendidikan formal. Rata-rata lama sekolah dihitung dari variabel pendidikan tertinggi yang ditamatkan dan tingkat pendidikan yang sedang dijalankan. Rata-rata lama sekolah merupakan salah satu indikator Indeks Pembangunan Manusia (IPM) atau *Human Development Index* (HDI).
3. Tingkat Pengangguran Terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Pengangguran terbuka terdiri dari
  - a. mereka yang tak punya pekerjaan dan mencari pekerjaan,
  - b. mereka yang tak punya pekerjaan dan mempersiapkan usaha,
  - c. mereka yang tak punya pekerjaan dan tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan, dan
  - d. mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja.

Bekerja adalah kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus) dalam seminggu yang lalu. Kegiatan tersebut termasuk pola kegiatan pekerja tak dibayar yang membantu dalam suatu usaha/kegiatan ekonomi.

4. Gini ratio merupakan koefisien yang didasarkan pada kurva lorenz, yaitu sebuah kurva pengeluaran kumulatif yang membandingkan distribusi dari suatu variabel tertentu (misalnya pendapatan) dengan distribusi uniform (seragam) yang mewakili persentase kumulatif penduduk. Gini rasio digunakan

untuk mengukur tingkat ketimpangan pendapatan secara menyeluruh. Koefisien gini berkisar antara 0 sampai 1. Apabila koefisien gini bernilai 0 berarti pemerataan sempurna, sedangkan apabila bernilai 1 berarti ketimpangan sempurna.

5. 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. Laju pertumbuhan ekonomi diukur dengan menggunakan laju pertumbuhan PDRB Atas Dasar Harga Konstan (ADHK). 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 2000.
6. PDRB perkapita adalah nilai PDRB dibagi jumlah penduduk dalam suatu wilayah per periode tertentu. PDRB per kapita atas dasar harga berlaku menunjukkan nilai PDRB per kepala atau per satu orang penduduk. PDRB per kapita atas dasar harga berlaku berguna untuk mengetahui pertumbuhan nyata ekonomi per kapita penduduk suatu negara.

### 3.3 Tahapan Penelitian

Tahap awal penelitian disesuaikan dengan tujuan pertama penelitian yaitu melakukan kajian estimator kurva regresi campuran nonparametrik spline dan kernel dalam model regresi nonparametrik aditif. Dalam kajian ini, satu pola data diasumsikan mengikuti kurva regresi spline, sedangkan pola data yang lainnya mengikuti kurva regresi kernel. Langkah-langkah yang dilakukan adalah sebagai berikut.

1. Menyajikan model regresi nonparametrik dimana variabel responnya adalah  $y_i$  dan variabel prediktornya adalah  $u_i, v_{1i}, v_{2i} \dots, v_{mi}$ .

$$y_i = f(u_i, \tilde{v}_i) + \varepsilon_i$$

dimana  $\tilde{v}_i = (v_{1i}, v_{2i} \dots, v_{mi})^T$ . Kurva  $f(u_i, \tilde{v}_i)$  diasumsikan aditif, sehingga

$$y_i = g(u_i) + h_1(v_{1i}) + h_2(v_{2i}) + \dots + h_m(v_{mi}) + \varepsilon_i$$

$$y_i = g(u_i) + \sum_{j=1}^m h_j(v_{ji}) + \varepsilon_i$$

Komponen kurva regresi  $g(u_i)$  diasumsikan mengikuti kurva regresi spline linier truncated, sedangkan komponen kurva regresi  $h_j(v_{ji})$  mengikuti kurva regresi kernel Nadaraya-Watson.

2. Menyajikan bentuk kurva regresi spline dalam bentuk vektor  $\tilde{g}(u)$  berdasarkan komponen-komponen  $g(u_i)$  sehingga terbentuk parameter  $\tilde{\theta}$ .

$$\tilde{g}(u) = \mathbf{G}(\tilde{\xi})\tilde{\theta}$$

3. Mendapatkan estimator komponen-komponen  $\hat{h}_j(v_{ji})$ .

$$\sum_{j=1}^m \hat{h}_j(v_j) = \mathbf{V}(\tilde{\phi})\tilde{y}$$

4. Menyajikan persamaan model regresi campuran nonparametrik berdasarkan hasil langkah 3 dan 4.

$$\tilde{y} = \mathbf{G}(\tilde{\xi})\tilde{\theta} + \mathbf{V}(\tilde{\phi})\tilde{y} + \tilde{\varepsilon}$$

5. Menyajikan bentuk optimasi untuk mendapatkan estimator  $\hat{\theta}$  dengan menggunakan metode *Maximum Likelihood Estimation* (MLE).

$$\underset{\tilde{\theta} \in R^{q+2}}{\text{Max}} \{L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})\}$$

6. Menyelesaikan optimasi menggunakan derivatif parsial.

$$\frac{\partial}{\partial \tilde{\theta}} (Q(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})) = 0$$

7. Mendapatkan estimator dari  $\tilde{g}(u)$  berdasarkan hasil estimasi dari  $\tilde{\theta}$ .

$$\hat{g}_{\tilde{\xi}}(u) = \mathbf{G}(\tilde{\xi})\hat{\theta}(\tilde{\xi}, \phi)$$

8. Mendapatkan estimator campuran dari  $\tilde{f}(u, \tilde{v})$  berdasarkan hasil estimator dari  $\tilde{g}(u)$  dan  $\sum_{j=1}^m \tilde{h}_j(v_j)$ .

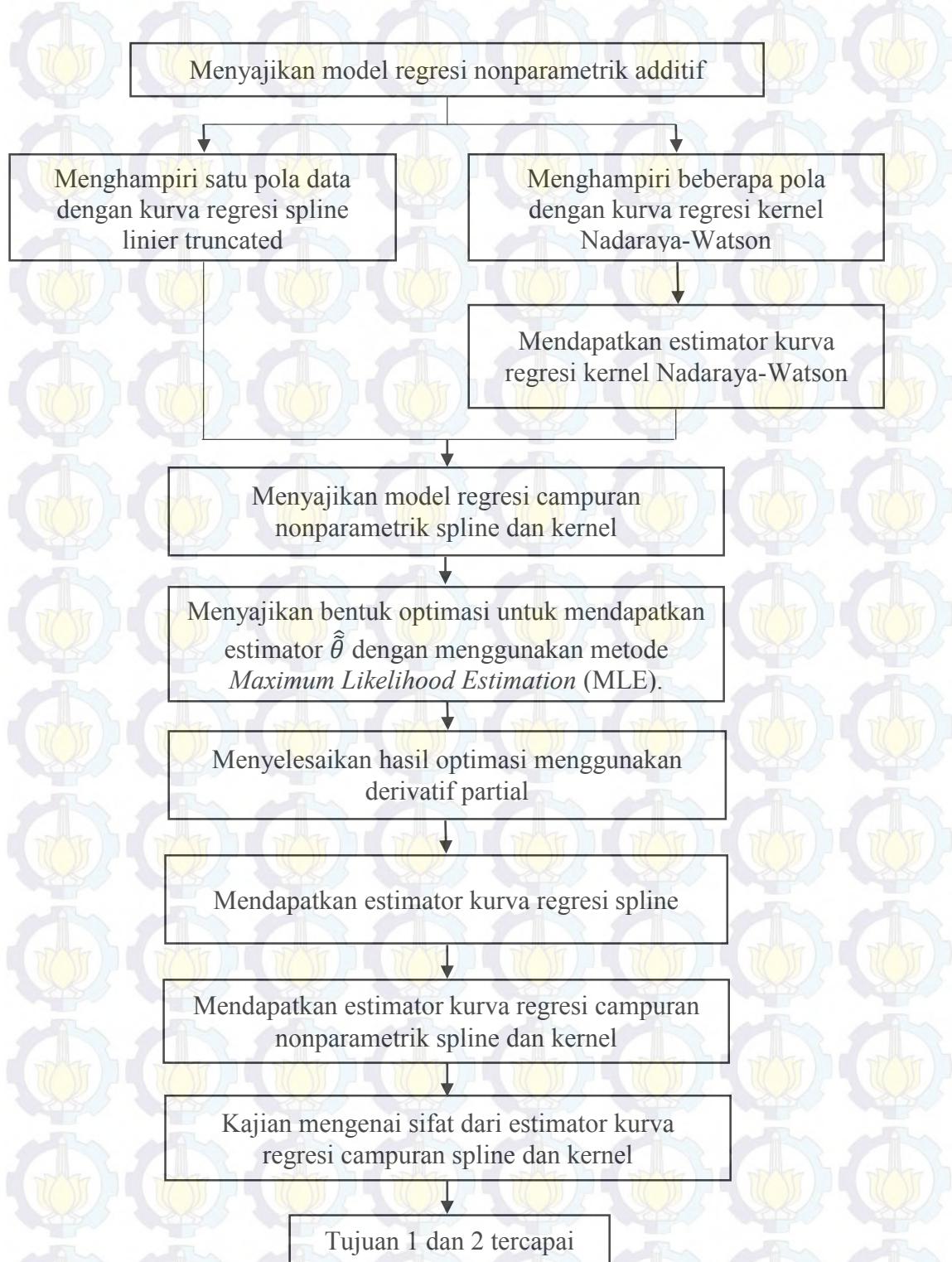
$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = \hat{g}_{\tilde{\xi}, \tilde{\phi}}(u_i, \tilde{v}_i) + \sum_{j=1}^m \hat{h}_j(v_{ji})$$

Tujuan kedua dari penelitian ini adalah melakukan kajian mengenai sifat estimator kurva regresi campuran spline dan kernel. Estimator yang dikaji adalah  $\hat{\theta}(\tilde{\xi}, \phi)$ ,  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i)$ ,  $\hat{g}_{\tilde{\xi}, \tilde{\phi}}(u_i, \tilde{v}_i)$  dan  $\sum_{j=1}^m \hat{h}_j(v_{ji})$ .

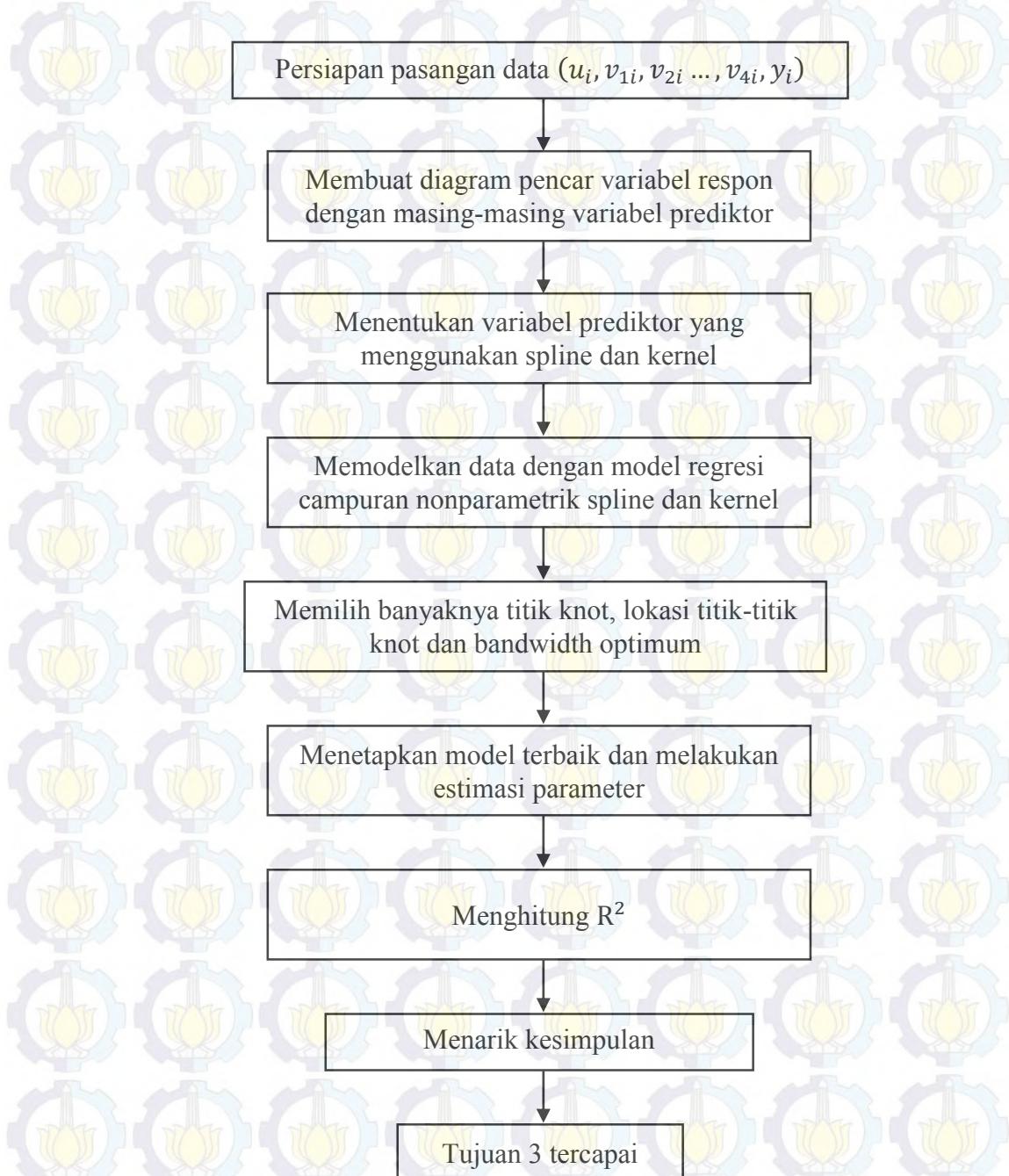
Selanjutnya tujuan ketiga adalah mengaplikasikan model regresi campuran nonparametrik spline linier truncated dan kernel pada data kemiskinan di Provinsi Papua. Langkah-langkahnya untuk menyelesaikan tujuan kedua tersebut adalah sebagai berikut.

1. Membuat diagram pencar (*scatter plot*) variabel respon dengan masing-masing variabel prediktor.
2. Menentukan variabel prediktor yang menggunakan kurva regresi spline dan yang menggunakan kurva regresi kernel.
3. Memodelkan data dengan estimator campuran nonparametrik spline dan kernel.
4. Memilih banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum secara simultan dan selanjutnya menetapkan model regresi campuran nonparametrik terbaik.
5. Melakukan estimasi model regresi campuran spline dan kernel.
6. Menghitung  $R^2$  sebagai bagian dari kriteria kebaikan model.
7. Mengambil kesimpulan.

Gambaran tahap analisis tujuan pertama dan kedua yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 3.1 dan 3.2.



Gambar 3.2 Diagram Alir Tahap Analisis untuk Tujuan Pertama dan Kedua



Gambar 3.3 Diagram Alir Tahap Analisis untuk Tujuan Ketiga

## BAB 4

# HASIL DAN PEMBAHASAN

Mengacu pada tujuan penelitian, bab ini akan melakukan kajian mengenai estimasi kurva regresi campuran nonparametrik spline dan kernel. Hasil kajian diaplikasikan pada data kemiskinan di Provinsi Papua tahun 2013. Bab ini terdiri dari lima subbab. Subbab 4.1 merupakan pengenalan bentuk model regresi campuran nonparametrik spline dan kernel. Subbab 4.2 berisi kajian estimasi kurva regresi campuran nonparametrik spline dan kernel yang dilanjutkan dengan Subbab 4.3 berupa kajian mengenai sifat dari estimator kurva regresi tersebut. Subbab 4.4 berisi pembahasan mengenai pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum. Selanjutnya pada Subbab 4.5, estimasi model regresi campuran nonparametrik spline dan kernel yang telah dibahas pada subbab-subbab sebelumnya, diaplikasikan pada data kemiskinan di Provinsi Papua.

### 4.1 Model Regresi Campuran Nonparametrik Spline dan Kernel

Diberikan data berpasangan  $(u_i, v_{1i}, v_{2i}, \dots, v_{mi}, y_i)$ ,  $i = 1, 2, \dots, n$  yang memiliki hubungan diasumsikan mengikuti model regresi nonparametrik. Variabel-variabel  $u_i, v_{1i}, v_{2i}, \dots, v_{mi}$  merupakan variabel-variabel prediktor dan variabel  $y_i$  merupakan variabel respon. Bentuk model regresi nonparametrik tersebut adalah

$$\begin{aligned} y_i &= f(u_i, v_{1i}, v_{2i}, \dots, v_{mi}) + \varepsilon_i \\ &= f(u_i, \tilde{v}_i) + \varepsilon_i, \end{aligned} \quad (4.1)$$

dimana  $\tilde{v}_i = (v_{1i}, v_{2i}, \dots, v_{mi})^T$ . Bentuk kurva regresi  $f(u_i, \tilde{v}_i)$  pada model (4.1) diasumsikan tidak diketahui dan hanya diketahui bahwa kurva tersebut *smooth* dalam arti kontinu dan differensiabel. Error random  $\varepsilon_i$  berdistribusi normal dengan  $E[\varepsilon_i] = 0$  dan  $\text{Var}[\varepsilon_i] = \sigma^2$ . Selain itu, kurva regresi  $f(u_i, \tilde{v}_i)$  diasumsikan bersifat aditif, sehingga dapat ditulis dalam bentuk

$$f(u_i, \tilde{v}_i) = g(u_i) + h_1(v_{1i}) + h_2(v_{2i}) + \dots + h_m(v_{mi}). \quad (4.2)$$

Bentuk pola hubungan variabel respon  $y_i$  dengan variabel prediktor  $u_i$  diasumsikan berubah-ubah pada sub-sub interval tertentu, sedangkan bentuk pola

hubungan variabel respon  $y_i$  dengan masing-masing variabel prediktor  $v_{1i}, v_{2i}, \dots, v_{mi}$  diasumsikan tidak diketahui atau tidak memiliki pola tertentu.

Secara teoritis, kurva regresi  $g(u_i)$  dapat dihampiri dengan kurva regresi spline, sedangkan masing-masing kurva regresi  $h_1(v_{1i}), h_2(v_{2i}), \dots, h_m(v_{mi})$  dapat dihampiri dengan kurva regresi kernel. Dengan demikian, kurva regresi  $f(u_i, \tilde{v}_i)$  disebut dengan kurva regresi campuran nonparametrik yang dikelompokkan menjadi dua komponen kurva regresi yaitu komponen kurva regresi spline dan komponen kurva regresi kernel. Selanjutnya, persamaan (4.2) dapat ditulis menjadi

$$f(u_i, \tilde{v}_i) = g(u_i) + \sum_{j=1}^m h_j(v_{ji}), \quad (4.3)$$

dimana

$$\sum_{j=1}^m h_j(v_{ji}) = h_1(v_{1i}) + h_2(v_{2i}) + \dots + h_m(v_{mi}).$$

Komponen kurva regresi  $g(u_i)$  merupakan komponen kurva regresi spline, sedangkan komponen kurva regresi  $\sum_{j=1}^m h_j(v_{ji})$  merupakan komponen kurva regresi kernel.

Komponen kurva regresi spline  $g(u_i)$  pada persamaan (4.3) didefinisikan oleh kurva regresi spline linier truncated

$$\begin{aligned} g(u_i) &= \beta_0 + \beta_1 u_i + \sum_{l=1}^q \lambda_l (u_i - \xi_l)_+ \\ &= \beta_0 + \beta_1 u_i + \lambda_1 (u_i - \xi_1)_+ + \dots + \lambda_q (u_i - \xi_q)_+, \end{aligned} \quad (4.4)$$

dimana

$$(u_i - \xi_l)_+ = \begin{cases} (u_i - \xi_l) & , u_i \geq \xi_l \\ 0 & , u_i < \xi_l. \end{cases}$$

Parameter-parameter  $\beta_0, \beta_1$  dan  $\lambda_1, \lambda_2, \dots, \lambda_q$  merupakan parameter-parameter yang tidak diketahui, sedangkan  $\xi_1, \xi_2, \dots, \xi_q$  merupakan titik-titik knot, dimana  $\xi_1 < \xi_2 < \dots < \xi_q$ . Selanjutnya, komponen kurva regresi kernel  $h_j(v_{ji})$  pada persamaan (4.3) didefinisikan oleh estimator kurva regresi kernel Nadaraya-Watson

$$\hat{h}_{\phi_j}(v_{ji}) = n^{-1} \sum_{i=1}^n W_{\phi_j i}(v_j) y_i. \quad (4.5)$$

Parameter  $\phi$  merupakan parameter bandwidth, sedangkan fungsi  $W_{\phi_j i}(v_j)$  merupakan fungsi pembobot yang didefinisikan oleh

$$W_{\phi_j i}(v_j) = \frac{K_{\phi_j}(v_j - v_{ji})}{n^{-1} \sum_{i=1}^n K_{\phi_j}(v_j - v_{ji})},$$

dimana  $K_{\phi_j}(v_j - v_{ji})$  adalah fungsi kernel

$$K_{\phi_j}(v_j - v_{ji}) = \frac{1}{\phi_j} K\left(\frac{v_j - v_{ji}}{\phi_j}\right).$$

Fungsi kernel  $K$  dapat berupa kernel uniform, kernel segitiga, kernel epanechnikov, kernel kuadrat, kernel triweight, kernel kosinus atau kernel Gaussian (Härdle, Müller, Sperlich, & Werwatz, 2004).

## 4.2 Estimasi Kurva Regresi

Jika kurva regresi spline linier truncated (4.4) berlaku untuk  $i = 1$  sampai dengan  $i = n$ , maka kumpulan persamaan-persamaan  $g(u_1), g(u_2), \dots, g(u_n)$  akan membentuk suatu persamaan vektor dan matriks

$$\begin{bmatrix} g(u_1) \\ g(u_2) \\ \vdots \\ g(u_n) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 u_1 + \lambda_1(u_1 - \xi_1)_+ + \dots + \lambda_q(u_1 - \xi_q)_+ \\ \beta_0 + \beta_1 u_2 + \lambda_1(u_2 - \xi_1)_+ + \dots + \lambda_q(u_2 - \xi_q)_+ \\ \vdots \\ \beta_0 + \beta_1 u_n + \lambda_1(u_n - \xi_1)_+ + \dots + \lambda_q(u_n - \xi_q)_+ \end{bmatrix},$$

selanjutnya menjadi

$$\begin{bmatrix} g(u_1) \\ g(u_2) \\ \vdots \\ g(u_n) \end{bmatrix} = \begin{bmatrix} 1 & u_1 & (u_1 - \xi_1)_+ & \dots & (u_1 - \xi_q)_+ \\ 1 & u_2 & (u_2 - \xi_1)_+ & \dots & (u_2 - \xi_q)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_n & (u_n - \xi_1)_+ & \dots & (u_n - \xi_q)_+ \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_q \end{bmatrix}. \quad (4.6)$$

Jika dinotasikan ke dalam bentuk notasi matriks, persamaan (4.6) menjadi persamaan kurva regresi spline linier truncated dalam bentuk vektor dan matriks yaitu

$$\tilde{g}(u) = \mathbf{G}(\tilde{\xi})\tilde{\theta}, \quad (4.7)$$

dimana

$$\tilde{g}(u) = \begin{bmatrix} g(u_1) \\ g(u_2) \\ \vdots \\ g(u_n) \end{bmatrix}, \quad \tilde{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_q \end{bmatrix},$$

$$\mathbf{G}(\tilde{\xi}) = \begin{bmatrix} 1 & u_1 & (u_1 - \xi_1)_+ & \cdots & (u_1 - \xi_q)_+ \\ 1 & u_2 & (u_2 - \xi_1)_+ & \cdots & (u_2 - \xi_q)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_n & (u_n - \xi_1)_+ & \cdots & (u_n - \xi_q)_+ \end{bmatrix}.$$

Vektor  $\tilde{g}(u)$  berukuran  $n \times 1$ , vektor  $\tilde{\theta}$  berukuran  $(q + 2) \times 1$ , dan matriks  $\mathbf{G}(\tilde{\xi})$  berukuran  $n \times (q + 2)$ .

Selanjutnya, setiap komponen kurva regresi kernel  $h_j(v_{ji})$  pada persamaan (4.3) diestimasi menggunakan estimator kernel Nadaraya-Watson (4.5)

$$\begin{aligned} \hat{h}_{j\phi_j}(v_{ji}) &= n^{-1} \sum_{i=1}^n W_{\phi_j i}(v_j) y_i \\ &= n^{-1} W_{\phi_j 1}(v_j) y_1 + n^{-1} W_{\phi_j 2}(v_j) y_2 + \cdots + n^{-1} W_{\phi_j n}(v_j) y_n. \end{aligned}$$

Persamaan tersebut berlaku pada  $i = 1$  sampai dengan  $i = n$ , sehingga kumpulan persamaan-persamaan  $\hat{h}_{j\phi_j}(v_{j1}), \hat{h}_{j\phi_j}(v_{j2}), \dots, \hat{h}_{j\phi_j}(v_{jn})$  dapat membentuk sebuah persamaan vektor dan matriks

$$\begin{aligned} \begin{bmatrix} \hat{h}_{j\phi_j}(v_{j1}) \\ \hat{h}_{j\phi_j}(v_{j2}) \\ \vdots \\ \hat{h}_{j\phi_j}(v_{jn}) \end{bmatrix} &= \begin{bmatrix} n^{-1} W_{\phi_j 1}(v_{j1}) y_1 + n^{-1} W_{\phi_j 2}(v_{j1}) y_2 + \cdots + n^{-1} W_{\phi_j n}(v_{j1}) y_n \\ n^{-1} W_{\phi_j 1}(v_{j2}) y_1 + n^{-1} W_{\phi_j 2}(v_{j2}) y_2 + \cdots + n^{-1} W_{\phi_j n}(v_{j2}) y_n \\ \vdots \\ n^{-1} W_{\phi_j 1}(v_{jn}) y_1 + n^{-1} W_{\phi_j 2}(v_{jn}) y_2 + \cdots + n^{-1} W_{\phi_j n}(v_{jn}) y_n \end{bmatrix} \\ &= \begin{bmatrix} n^{-1} W_{\phi_j 1}(v_{j1}) & n^{-1} W_{\phi_j 2}(v_{j1}) & \cdots & n^{-1} W_{\phi_j n}(v_{j1}) \\ n^{-1} W_{\phi_j 1}(v_{j2}) & n^{-1} W_{\phi_j 2}(v_{j2}) & \cdots & n^{-1} W_{\phi_j n}(v_{j2}) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\phi_j 1}(v_{jn}) & n^{-1} W_{\phi_j 2}(v_{jn}) & \cdots & n^{-1} W_{\phi_j n}(v_{jn}) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \end{aligned}$$

Selanjutnya dapat ditulis dalam bentuk notasi matriks

$$\hat{h}_{j\phi_j}(v_j) = \mathbf{V}_j(\boldsymbol{\phi}_j)\tilde{y}, \quad (4.8)$$

dimana

$$\begin{aligned} \hat{h}_{j\phi_j}(v_j) &= \begin{bmatrix} \hat{h}_{j\phi_j}(v_{j1}) \\ \hat{h}_{j\phi_j}(v_{j2}) \\ \vdots \\ \hat{h}_{j\phi_j}(v_{jn}) \end{bmatrix}, \quad \tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \\ \mathbf{V}_j(\boldsymbol{\phi}_j) &= \begin{bmatrix} n^{-1}W_{\phi_j1}(v_{j1}) & n^{-1}W_{\phi_j2}(v_{j1}) & \cdots & n^{-1}W_{\phi_jn}(v_{j1}) \\ n^{-1}W_{\phi_j1}(v_{j2}) & n^{-1}W_{\phi_j2}(v_{j2}) & \cdots & n^{-1}W_{\phi_jn}(v_{j2}) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_{\phi_j1}(v_{jn}) & n^{-1}W_{\phi_j2}(v_{jn}) & \cdots & n^{-1}W_{\phi_jn}(v_{jn}) \end{bmatrix}. \end{aligned}$$

Vektor  $\hat{h}_{j\phi_j}(v_j)$  berukuran  $n \times 1$ , vektor  $\tilde{y}$  berukuran  $n \times 1$ , dan matriks  $\mathbf{V}_j(\boldsymbol{\phi}_j)$  berukuran  $n \times n$ . Berdasarkan persamaan (4.8), maka estimator untuk komponen kurva regresi kernel  $\sum_{j=1}^m h_j(v_{ji})$  pada persamaan (4.3) akan menjadi

$$\begin{aligned} \sum_{j=1}^m \hat{h}_{j\phi_j}(v_j) &= \sum_{j=1}^m \mathbf{V}_j(\boldsymbol{\phi}_j)\tilde{y} \\ &= \mathbf{V}_1(\boldsymbol{\phi}_1)\tilde{y} + \mathbf{V}_2(\boldsymbol{\phi}_2)\tilde{y} + \cdots + \mathbf{V}_m(\boldsymbol{\phi}_m)\tilde{y} \\ &= \{\mathbf{V}_1(\boldsymbol{\phi}_1) + \mathbf{V}_2(\boldsymbol{\phi}_2) + \cdots + \mathbf{V}_m(\boldsymbol{\phi}_m)\}\tilde{y} \\ &= \mathbf{V}(\tilde{\boldsymbol{\phi}})\tilde{y}, \end{aligned} \quad (4.9)$$

dimana

$$\begin{aligned} \mathbf{V}(\tilde{\boldsymbol{\phi}}) &= \mathbf{V}_1(\boldsymbol{\phi}_1) + \mathbf{V}_2(\boldsymbol{\phi}_2) + \cdots + \mathbf{V}_m(\boldsymbol{\phi}_m) \\ &= \sum_{j=1}^m \mathbf{V}_j(\boldsymbol{\phi}_j) \\ &= \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\phi_j1}(v_{j1}) & n^{-1} \sum_{i=1}^n W_{\phi_j2}(v_{j1}) & \cdots & n^{-1} \sum_{i=1}^n W_{\phi_jn}(v_{j1}) \\ n^{-1} \sum_{i=1}^n W_{\phi_j1}(v_{j2}) & n^{-1} \sum_{i=1}^n W_{\phi_j2}(v_{j2}) & \cdots & n^{-1} \sum_{i=1}^n W_{\phi_jn}(v_{j2}) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} \sum_{i=1}^n W_{\phi_j1}(v_{jn}) & n^{-1} \sum_{i=1}^n W_{\phi_j2}(v_{jn}) & \cdots & n^{-1} \sum_{i=1}^n W_{\phi_jn}(v_{jn}) \end{bmatrix}. \end{aligned}$$

Matriks  $\mathbf{V}(\tilde{\boldsymbol{\phi}})$  berukuran  $n \times n$ .

Jika kurva regresi spline linier truncated  $\tilde{g}(u)$  diberikan oleh persamaan (4.7) dan estimator untuk komponen kurva regresi kernel  $\sum_{j=1}^m h_j(v_{ji})$  diberikan oleh persamaan (4.9), maka model regresi campuran nonparametrik spline dan kernel (4.1) dapat disajikan dalam bentuk vektor dan matriks

$$\begin{aligned}\tilde{y} &= \tilde{g}(u) + \sum_{j=1}^m \hat{h}_{j\phi_j}(v_j) + \tilde{\varepsilon} \\ &= \mathbf{G}(\tilde{\xi})\tilde{\theta} + \mathbf{V}(\tilde{\phi})\tilde{y} + \tilde{\varepsilon},\end{aligned}\quad (4.10)$$

dimana

$$\tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Vektor  $\tilde{\varepsilon}$  merupakan vektor error berukuran  $n \times 1$ . Berdasarkan persamaan (4.10), diperoleh vektor error

$$\begin{aligned}\tilde{\varepsilon} &= \tilde{y} - \mathbf{V}(\tilde{\phi})\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta} \\ &= (\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}.\end{aligned}$$

Matriks  $\mathbf{I}$  merupakan matriks identitas berukuran  $n \times n$ . Selanjutnya, jumlah kuadrat error adalah

$$\|\tilde{\varepsilon}\|^2 = \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2. \quad (4.11)$$

Error random  $\tilde{\varepsilon}$  berdistribusi multivariat normal dengan  $E[\tilde{\varepsilon}] = 0$  dan  $E[\tilde{\varepsilon}\tilde{\varepsilon}^T] = \sigma^2\mathbf{I}$ , sehingga fungsi likelihood  $L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})$  diberikan oleh

$$\begin{aligned}L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \varepsilon_i^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|\tilde{\varepsilon}\|^2\right).\end{aligned}$$

Jika jumlah kuadrat error diberikan oleh persamaan (4.11), maka

$$L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2\right).$$

Berdasarkan metode MLE, estimator untuk parameter  $\tilde{\theta}$  diperoleh dari optimasi

$$\underset{\tilde{\theta} \in R^{q+2}}{\text{Max}} \{L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})\} = \underset{\tilde{\theta} \in R^{q+2}}{\text{Max}} \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2\right) \right\}.$$

Jika fungsi likelihood  $L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})$  ditransformasi ke bentuk logaritma natural

$$\begin{aligned} l(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi}) &= \ln L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi}) \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2, \end{aligned}$$

maka optimasi tersebut akan menjadi

$$\underset{\tilde{\theta} \in R^{q+2}}{\text{Max}} \{l(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})\} = \underset{\tilde{\theta} \in R^{q+2}}{\text{Max}} \left\{ -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2 \right\}.$$

Optimasi akan maksimum ketika komponen

$$Q(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi}) = \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2$$

mempunyai nilai yang minimum, sehingga berlaku persamaan

$$\underset{\tilde{\theta} \in R^{q+2}}{\text{Max}} \{L(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})\} = \underset{\tilde{\theta} \in R^{q+2}}{\text{Min}} \left\{ \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2 \right\}.$$

Untuk mendapatkan estimator dari  $\tilde{\theta}$ , maka perlu dilakukan derivatif parsial terhadap

$$\begin{aligned} Q(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi}) &= \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}\|^2 \\ &= [(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}]^T [(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}] \\ &= [\tilde{y}^T (\mathbf{I} - \mathbf{V}(\tilde{\phi}))^T - \tilde{\theta}^T (\mathbf{G}(\tilde{\xi}))^T] [(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} - \mathbf{G}(\tilde{\xi})\tilde{\theta}] \\ &= \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y}\|^2 - 2\tilde{\theta}^T (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} + \\ &\quad \tilde{\theta}^T (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi})\tilde{\theta}. \end{aligned}$$

Selanjutnya, derivatif parsial terhadap  $Q(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})$  disamakan dengan 0,

$$\begin{aligned} \frac{\partial}{\partial \tilde{\theta}} [Q(\tilde{\theta}, \sigma^2 | \tilde{\phi}, \tilde{\xi})] &= 0 \\ \frac{\partial}{\partial \tilde{\theta}} \left[ \|(\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y}\|^2 - 2\tilde{\theta}^T (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} + \tilde{\theta}^T (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi})\tilde{\theta} \right] &= 0. \end{aligned}$$

Derivatif parsial akan menghasilkan persamaan normal

$$(\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi})\tilde{\theta} = (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y}.$$

Sehingga estimasi untuk  $\tilde{\theta}$  diberikan oleh

$$\begin{aligned} \hat{\theta}(\tilde{\xi}, \tilde{\phi}) &= [(\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi})]^{-1} (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi}))\tilde{y} \\ &= \mathbf{B}(\tilde{\xi}, \tilde{\phi})\tilde{y}, \end{aligned} \tag{4.12}$$

dimana

$$\mathbf{B}(\tilde{\xi}, \tilde{\phi}) = \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi})),$$

$$\hat{\theta}(\tilde{\xi}, \tilde{\phi}) = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \vdots \\ \hat{\lambda}_q \end{bmatrix}.$$

Matriks  $\mathbf{B}(\tilde{\xi}, \tilde{\phi})$  berukuran  $n \times n$  dan vektor  $\hat{\theta}(\tilde{\xi}, \tilde{\phi})$  berukuran  $(q + 2) \times 1$ .

Mengingat persamaan (4.12) dan sifat invariant dari MLE, maka estimator kurva regresi spline linier truncated (4.7) adalah

$$\begin{aligned} \hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) &= \mathbf{G}(\tilde{\xi}) \hat{\theta}(\tilde{\xi}, \tilde{\phi}) \\ &= \mathbf{G}(\tilde{\xi}) \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi})) \tilde{y} \\ &= \mathbf{S}(\tilde{\xi}, \tilde{\phi}) \tilde{y}, \end{aligned} \quad (4.13)$$

dimana

$$\mathbf{S}(\tilde{\xi}, \tilde{\phi}) = \mathbf{G}(\tilde{\xi}) \left[ (\mathbf{G}(\tilde{\xi}))^T \mathbf{G}(\tilde{\xi}) \right]^{-1} (\mathbf{G}(\tilde{\xi}))^T (\mathbf{I} - \mathbf{V}(\tilde{\phi})).$$

Matriks  $\mathbf{S}(\tilde{\xi}, \tilde{\phi})$  berukuran  $n \times n$ .

Berdasarkan persamaan (4.9) dan (4.13), maka estimator kurva regresi campuran nonparametrik spline dan kernel (4.3) adalah

$$\begin{aligned} \hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) &= \hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) + \sum_{j=1}^m \hat{h}_{j\phi_j}(v_j) \\ &= \mathbf{S}(\tilde{\xi}, \tilde{\phi}) \tilde{y} + \mathbf{V}(\tilde{\phi}) \tilde{y} \\ &= (\mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})) \tilde{y} \\ &= \mathbf{Z}(\tilde{\xi}, \tilde{\phi}) \tilde{y}, \end{aligned}$$

dimana

$$\mathbf{Z}(\tilde{\xi}, \tilde{\phi}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi}).$$

Matriks  $\mathbf{Z}(\tilde{\xi}, \tilde{\phi})$  berukuran  $n \times n$ .

### 4.3 Sifat Estimator Kurva Regresi

Berikut ini akan diselidiki sifat dari estimator  $\hat{\theta}(\tilde{\xi}, \tilde{\phi})$ , estimator kurva regresi spline  $\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$ , estimator kurva regresi kernel  $\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)$ , dan estimator kurva regresi campuran nonparametrik spline dan kernel  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$ . Sifat ini diselidiki dari nilai harapan estimator tersebut. Nilai harapan dari estimator  $\hat{\theta}(\tilde{\xi}, \tilde{\phi})$  adalah

$$\begin{aligned}
E[\hat{\theta}(\tilde{\xi}, \tilde{\phi})] &= E\left[\left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T (I - V(\tilde{\phi})) \tilde{y}\right] \\
&= \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T (I - V(\tilde{\phi})) E[\tilde{y}] \\
&= \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T (I - V(\tilde{\phi})) \left(\tilde{g}(u) + \sum_{j=1}^m \tilde{h}_j(v_j)\right) \\
&= \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T (I - V(\tilde{\phi})) \left(G(\tilde{\xi})\tilde{\theta} + \sum_{j=1}^m \tilde{h}_j(v_j)\right) \\
&= \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\tilde{\theta} + \\
&\quad \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T \sum_{j=1}^m \tilde{h}_j(v_j) - \\
&\quad \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T V(\tilde{\phi})G(\tilde{\xi})\tilde{\theta} - \\
&\quad \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T V(\tilde{\phi}) \sum_{j=1}^m \tilde{h}_j(v_j) \\
&= \tilde{\theta} + \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T \sum_{j=1}^m \tilde{h}_j(v_j) - \\
&\quad \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T V(\tilde{\phi})G(\tilde{\xi})\tilde{\theta} - \\
&\quad \left[\left(G(\tilde{\xi})\right)^T G(\tilde{\xi})\right]^{-1}\left(G(\tilde{\xi})\right)^T V(\tilde{\phi}) \sum_{j=1}^m \tilde{h}_j(v_j) \\
&\neq \tilde{\theta}.
\end{aligned}$$

Nilai harapan dari estimator kurva regresi spline  $\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$  adalah

$$\begin{aligned}
 E[\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})] &= E[S(\tilde{\xi}, \tilde{\phi})\tilde{y}] \\
 &= S(\tilde{\xi}, \tilde{\phi})E[\tilde{y}] \\
 &= S(\tilde{\xi}, \tilde{\phi}) \left( \tilde{g}(u) + \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) \right) \\
 &= S(\tilde{\xi}, \tilde{\phi})\tilde{g}(u) + S(\tilde{\xi}, \tilde{\phi}) \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) \\
 &\neq \tilde{g}(u).
 \end{aligned}$$

Nilai harapan dari estimator kurva regresi kernel  $\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)$  adalah

$$\begin{aligned}
 E\left[\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)\right] &= E[V(\tilde{\phi})\tilde{y}] \\
 &= V(\tilde{\phi})E[\tilde{y}] \\
 &= V(\tilde{\phi}) \left( \tilde{g}(u) + \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) \right) \\
 &= V(\tilde{\phi})\tilde{g}(u) + V(\tilde{\phi}) \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) \\
 &\neq \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j).
 \end{aligned}$$

Nilai harapan dari estimator kurva regresi campuran nonparametrik spline dan kernel  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$  adalah

$$\begin{aligned}
 E[\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})] &= E\left[\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) + \sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)\right] \\
 &= E[(S(\tilde{\xi}, \tilde{\phi}) + V(\tilde{\phi}))\tilde{y}] \\
 &= (S(\tilde{\xi}, \tilde{\phi}) + V(\tilde{\phi}))E[\tilde{y}] \\
 &= (S(\tilde{\xi}, \tilde{\phi}) + V(\tilde{\phi})) \left( \tilde{g}(u) + \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) \right)
 \end{aligned}$$

$$\begin{aligned}
&= (\mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})) \tilde{g}(u) + (\mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})) \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) \\
&\neq \tilde{g}(u) + \sum_{j=1}^m \tilde{h}_{j\phi_j}(v_j) = \tilde{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}).
\end{aligned}$$

Estimator-estimator  $\hat{\theta}(\tilde{\xi}, \tilde{\phi})$ ,  $\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$ ,  $\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)$  dan  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$  seperti pada umumnya estimator kurva regresi nonparametrik yang lain, bersifat bias. Walaupun demikian estimator-estimator tersebut masih merupakan kelas estimator linier dalam observasi. Hal ini dapat diketahui dari pembahasan 4.2 yang menghasilkan

$$\begin{aligned}
\hat{\theta}(\tilde{\xi}, \tilde{\phi}) &= \mathbf{B}(\tilde{\xi}, \tilde{\phi})\tilde{y}, \quad \hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi})\tilde{y}, \\
\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j) &= \mathbf{V}(\tilde{\phi})\tilde{y}, \quad \hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{Z}(\tilde{\xi}, \tilde{\phi})\tilde{y}.
\end{aligned}$$

Terlihat bahwa estimator-estimator  $\hat{\theta}(\tilde{\xi}, \tilde{\phi})$ ,  $\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$ ,  $\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)$ , dan  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$  merupakan kelas estimator linier dalam observasi  $\tilde{y}$ .

#### 4.4 Pemilihan Titik-titik Knot dan Bandwidth Secara Simultan

Estimator kurva regresi campuran nonparametrik spline dan kernel  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$  sangat tergantung pada banyaknya titik knot, lokasi titik-titik knot dan bandwidth. Untuk memperoleh estimator kurva regresi campuran nonparametrik spline dan kernel terbaik, diperlukan pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum. Salah satu metode yang digunakan untuk melakukan pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum adalah metode *Generalized Cross Validation* atau GCV (Wahba, 1990).

$$GCV(\tilde{\xi}, \tilde{\phi}) = \frac{\text{MSE}(\tilde{\xi}, \tilde{\phi})}{\left(n^{-1}\text{trace}(\mathbf{I} - \mathbf{Z}(\tilde{\xi}, \tilde{\phi}))\right)^2}, \quad (4.14)$$

dimana

$$\text{MSE}(\tilde{\xi}, \tilde{\phi}) = n^{-1} \sum_{i=1}^n \left(y_i - \hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i)\right)^2.$$

Banyaknya titik knot optimum dan lokasi titik-titik knot optimum  $\tilde{\xi}_{(opt)} = (\xi_{1(opt)}, \xi_{2(opt)}, \dots, \xi_{q(opt)})^T$  serta bandwidth optimum  $\tilde{\phi}_{(opt)} = (\phi_{1(opt)}, \phi_{2(opt)}, \dots, \phi_{m(opt)})^T$  diperoleh dari optimasi

$$GCV(\tilde{\xi}_{(opt)}, \tilde{\phi}_{(opt)}) = \underset{\tilde{\xi}, \tilde{\phi}}{\text{Min}} \{GCV(\tilde{\xi}, \tilde{\phi})\}. \quad (4.15)$$

Penyelesaian optimasi dengan meminimumkan fungsi  $GCV(\tilde{\xi}, \tilde{\phi})$  pada persamaan (4.15) bisa diselesaikan salah satunya dengan memanfaatkan metode optimasi model fungsi non linier dengan pembatas (*constrained nonlinear regression optimization*) prosedur nlminb pada program R. Prosedur ini meminimumkan fungsi GCV untuk mendapatkan lokasi titik-titik knot dan bandwidth-bandwidth optimum secara simultan. Sebelum melakukan optimasi, titik-titik knot  $\tilde{\xi}$  dikombinasikan terlebih dahulu, dimana pada kombinasi tersebut  $\xi_1 < \xi_2 < \dots < \xi_q$ . Selanjutnya, pada setiap kombinasi dilakukan optimasi GCV terhadap bandwidth  $\tilde{\phi}$ . Fungsi nlminb disajikan pada Lampiran 2, sedangkan prosedur pemanggilan nlminb dalam *software R* disajikan pada Lampiran 3.

## 4.5 Aplikasi pada Data Kemiskinan

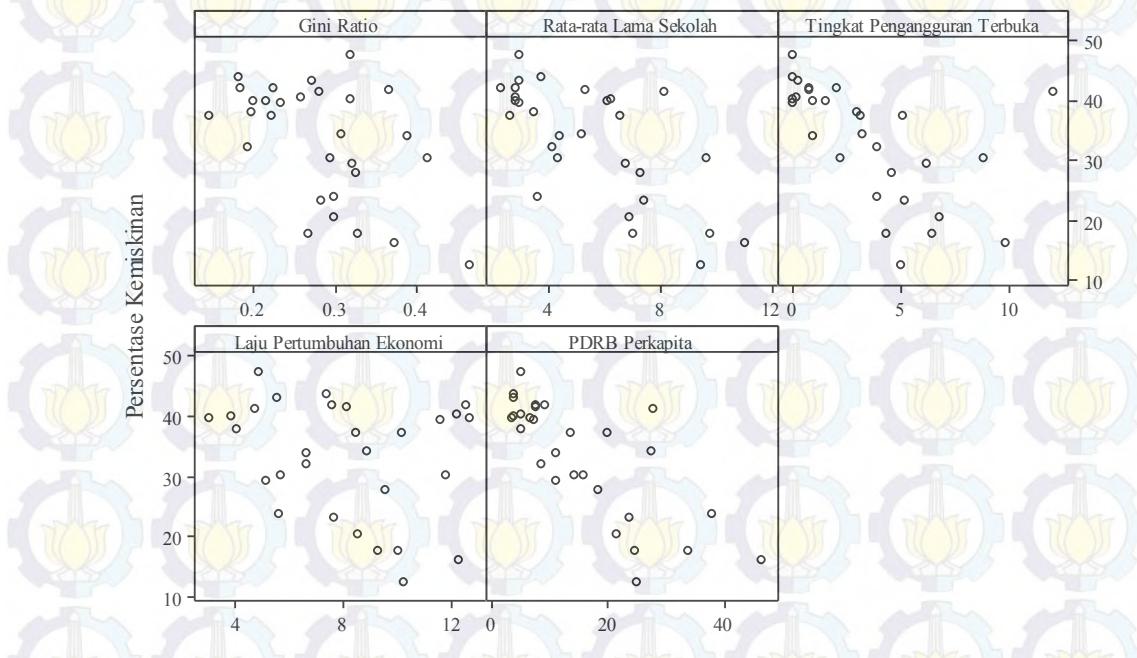
Model regresi campuran nonparametrik spline dan kernel pada pembahasan 4.1 sampai dengan 4.4 akan diaplikasikan pada data kemiskinan kabupaten/kota se-Provinsi Papua. Banyaknya variabel yang digunakan adalah enam variabel, yaitu satu variabel respon dan lima variabel prediktor. Data yang dijadikan sebagai variabel respon adalah persentase penduduk miskin (persen), sedangkan sebagai variabel prediktor adalah gini ratio, rata-rata lama sekolah (tahun), tingkat pengangguran terbuka atau TPT (persen), laju pertumbuhan ekonomi dan PDRB perkapita (juta rupiah). Jumlah unit observasi sebanyak 29 kabupaten/kota.

### 4.5.1 Eksplorasi Data

Tahap awal sebelum melakukan pemodelan regresi adalah mendapatkan informasi mengenai bentuk pola hubungan antara variabel respon dengan setiap variabel prediktor. Informasi tersebut digunakan untuk menentukan jenis kurva

regresi yang sesuai dalam menghampiri masing-masing pola data. Informasi mengenai bentuk pola hubungan antara variabel respon dan variabel prediktor dapat diperoleh dari plot sebaran data yang disajikan dalam bentuk diagram pencar.

Gambar 4.1 merupakan diagram pencar antara variabel respon dengan masing-masing variabel prediktor. Pada diagram pencar tersebut terlihat bahwa secara umum bentuk pola hubungan antara variabel respon dan variabel prediktor tidak diketahui atau tidak mengikuti pola tertentu. Namun jika diperhatikan dengan lebih seksama, pada pola data pada diagram pencar antara persentase penduduk miskin dengan PDRB perkapita, akan terlihat bahwa pola perilaku datanya berubah-ubah pada sub-sub interval tertentu. Dengan demikian, bentuk pola hubungan antara persentase penduduk miskin dengan PDRB perkapita dihampiri dengan kurva regresi spline, sedangkan bentuk pola hubungan antara persentase kemiskinan dengan empat variabel lainnya (gini ratio, rata-rata lama sekolah, tingkat pengangguran terbuka dan laju pertumbuhan ekonomi) masing-masing dihampiri dengan kurva regresi kernel.



Gambar 4.1 Diagram Pencar Variabel Respon dan Prediktor

Selanjutnya, jika bentuk pola hubungan variabel respon dan variabel-variabel prediktor tersebut didekati dengan model regresi campuran nonparametrik

spline dan kernel (4.3) dimana variabel PDRB perkapita dihampiri dengan kurva regresi kernel dan variabel gini ratio, rata-rata lama sekolah, tingkat pengangguran terbuka dan laju pertumbuhan ekonomi dihampiri dengan kurva regresi spline, maka variabel-variabel yang digunakan tersebut dapat dinotasikan menjadi

$y$  = persentase penduduk miskin,

$u$  = PDRB perkapita,

$v_1$  = gini ratio,

$v_2$  = rata-rata lama sekolah,

$v_3$  = tingkat pengangguran terbuka,

$v_4$  = laju pertumbuhan ekonomi.

Data untuk variabel-variabel tersebut disajikan pada Lampiran 1. Ringkasan data disajikan pada Tabel 4.1.

Tabel 4.1 Statistik Deskriptif Variabel Respon dan Prediktor

Statistik Deskriptif	$y$	$u$	$v_1$	$v_2$	$v_3$	$v_4$
Min	12,33	3,5871	0,15	2,30	0,00	3,05
Max	47,52	46,5405	0,47	11,07	11,98	12,69
Range	35,19	42,9534	0,32	8,77	11,98	9,64
Rata-rata	33,2124	15,5921	0,2822	5,4914	3,4617	8,1034
Standar Deviasi	9,6856	11,4664	0,0765	2,5277	3,1856	2,8926

#### 4.5.2 Model Regresi Campuran Nonparametrik

Berdasarkan eksplorasi data pada pembahasan 4.5.1, maka model regresi campuran nonparametrik spline dan kernel dapat diaplikasikan pada data, dimana dalam model tersebut terdapat satu kurva regresi spline dan empat kurva regresi kernel. Dengan demikian, berdasarkan pasangan data yang diberikan  $(u_i, v_{1i}, v_{2i}, v_{3i}, v_{4i}, y_i)$ ,  $i = 1, 2, \dots, 29$ , maka model regresi campuran nonparametrik spline dan kernel tersebut adalah

$$y_i = f(u_i, \tilde{v}_i) + \varepsilon_i,$$

dimana  $\tilde{v}_i = (v_{1i}, v_{2i}, v_{3i}, v_{4i})^T$ , sehingga dalam bentuk aditif

$$y_i = g(u_i) + h_1(v_{1i}) + h_2(v_{2i}) + h_3(v_{3i}) + h_4(v_{4i}) + \varepsilon_i,$$

selanjutnya

$$\begin{aligned}
y_i = & \beta_0 + \beta_1 u_i + \lambda_1(u_i - \xi_1)_+ + \cdots + \lambda_q(u_i - \xi_q)_+ + \\
& \sum_{i=1}^{29} \frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right) y_i + \sum_{i=1}^{29} \frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right) y_i + \\
& \sum_{i=1}^{29} \frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right) y_i + \sum_{i=1}^{29} \frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right) y_i + \varepsilon_i .
\end{aligned}$$

Dalam model regresi tersebut, terdapat  $q$  titik knot  $(\xi_1, \xi_2, \dots, \xi_q)$  dan  $m = 4$  bandwidth  $(\phi_1, \phi_2, \phi_3, \phi_4)$ . Dalam penelitian ini, banyaknya titik knot dibatasi hingga  $q = 3$  titik knot  $(\xi_1, \xi_2, \xi_3)$ .

#### 4.5.3 Fungsi Kernel, Titik Knot dan Bandwidth Optimum

Dalam pemodelan regresi campuran nonparametrik spline dan kernel, perlu dilakukan penentuan jenis fungsi kernel  $K$  dan pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum. Fungsi kernel yang digunakan dalam penelitian ini adalah fungsi kernel Gaussian (2.12), sedangkan pemilihan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum dilakukan secara simultan dengan meminimumkan fungsi GCV (4.14).

Jika model regresi campuran nonparametrik spline dan kernel memiliki banyaknya titik knot yang dibatasi sampai dengan tiga titik knot, maka terdapat tiga kemungkinan model yang bisa dibentuk, yaitu model regresi campuran spline dan kernel dengan satu titik knot, model regresi campuran spline dan kernel dengan dua titik knot, dan model regresi campuran spline dan kernel dengan tiga titik knot. Berikut ini adalah pemilihan lokasi titik-titik knot dan bandwidth optimum ketiga bentuk model tersebut berdasarkan metode GCV minimum. Selanjutnya, GCV minimum setiap model dibandingkan untuk mendapatkan banyaknya titik knot yang optimum.

##### 4.5.3.1 Model dengan Satu Titik Knot

Model regresi campuran nonparametrik spline dan kernel dengan satu titik knot dan empat bandwidth  $\phi_1, \phi_2, \phi_3$  dan  $\phi_4$  adalah

$$\begin{aligned}
y_i = & \beta_0 + \beta_1 u_i + \lambda_1 (u_i - \xi_1)_+ + \sum_{i=1}^{29} \frac{\frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right)} y_i + \\
& \sum_{i=1}^{29} \frac{\frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right)} y_i + \\
& \sum_{i=1}^{29} \frac{\frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right)} y_i + \varepsilon_i .
\end{aligned}$$

Syntax R yang digunakan untuk mendapatkan GCV model dengan satu titik knot disajikan pada Lampiran 4. Lokasi titik-titik knot, bandwidth dan GCV yang diperoleh disajikan pada Tabel 4.2.

Tabel 4.2 GCV pada Model dengan Satu Titik Knot

No	Titik Knot		Bandwidth				GCV
	$\xi_1$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$		
1	14,0186	0,1139	0,0696	0,0389	0,0077	22,0305	
2	16,0129	0,1160	0,0686	0,0388	0,0078	21,9732	
3	18,0839	0,1155	0,0693	0,0386	0,0077	21,8351	
4	20,9986	0,1153	0,0712	0,0387	0,0083	21,5841	
<b>5</b>	<b>24,1434</b>	<b>0,1182</b>	<b>0,0751</b>	<b>0,0395</b>	<b>0,0077</b>	<b>21,2932</b>	
6	25,0638	0,1192	0,0765	0,0395	0,0078	21,3468	
7	26,1376	0,1165	0,0769	0,0397	0,0076	21,7099	
8	27,0581	0,1143	0,0770	0,0398	0,0076	21,9926	
9	28,0552	0,1123	0,0768	0,0398	0,0075	22,1998	
10	30,9699	0,1125	0,0781	0,0398	0,0078	21,9427	

Tabel 4.2 menunjukkan bahwa nilai GCV minimum yang dihasilkan adalah 21,2932. GCV tersebut diperoleh ketika lokasi titik knot ( $\xi_1$ ) berada pada 24,1434 dan bandwidth ( $\phi_1, \phi_2, \phi_3, \phi_4$ ) sama dengan 0,1182, 0,0751, 0,0395 dan 0,0077.

#### 4.5.3.2 Model dengan Dua Titik Knot

Model regresi campuran nonparametrik spline dan kernel dengan dua titik knot dan empat bandwidth  $\phi_1, \phi_2, \phi_3$  dan  $\phi_4$  adalah sebagai

$$y_i = \beta_0 + \beta_1 u_i + \lambda_1 (u_i - \xi_1)_+ + \lambda_2 (u_i - \xi_2)_+ +$$

$$\begin{aligned}
& \sum_{i=1}^{29} \frac{\frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right)} y_i + \\
& \sum_{i=1}^{29} \frac{\frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right)} y_i + \varepsilon_i .
\end{aligned}$$

Syntax R yang digunakan untuk mendapatkan GCV model dengan dua titik knot disajikan pada Lampiran 5. Lokasi titik-titik knot, bandwidth dan GCV yang diperoleh disajikan pada tabel 4.3.

Tabel 4.3 GCV pada Model dengan Dua Titik Knot

No	Titik Knot		Bandwidth				GCV
	$\xi_1$	$\xi_2$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	
1	12,6984	21,8098	0,1150	0,0706	0,0388	0,0077	22,8073
2	16,6033	23,1114	0,1139	0,0726	0,0395	0,0029	22,4439
3	17,9049	27,0162	0,1135	0,0659	0,0383	0,0077	22,6556
4	20,5081	28,3178	0,1110	0,0617	0,0369	0,0078	21,9038
5	21,8098	29,6195	0,1129	0,0601	0,0373	0,0078	21,4083
6	23,1114	28,3178	0,1145	0,0557	0,0369	0,0077	19,5870
7	<b>25,7146</b>	<b>28,3178</b>	<b>0,1387</b>	<b>0,0540</b>	<b>0,0361</b>	<b>0,0076</b>	<b>17,1844</b>
8	25,7146	30,9211	0,1249	0,0671	0,0393	0,0021	21,5130
9	27,0162	34,8259	0,1109	0,0793	0,0401	0,0077	22,6506
10	28,3178	34,8259	0,1071	0,0812	0,0399	0,0078	22,4255

Tabel 4.3 menunjukkan bahwa nilai GCV minimum yang dihasilkan adalah 17,1844.

Nilai tersebut diperoleh ketika lokasi titik-titik knot ( $\xi_1, \xi_2$ ) berada pada 25,7146 dan 28,3178 dan bandwidth ( $\phi_1, \phi_2, \phi_3, \phi_4$ ) sama dengan 0,1387, 0,0540, 0,0361 dan 0,0076.

#### 4.5.3.3 Model dengan Tiga Titik Knot

Model regresi campuran nonparametrik spline dan kernel dengan tiga titik knot dan empat bandwidth  $\phi_1, \phi_2, \phi_3$  dan  $\phi_4$  adalah

$$\begin{aligned}
y_i = & \beta_0 + \beta_1 u_i + \lambda_1(u_i - \xi_1)_+ + \lambda_2(u_i - \xi_2)_+ + \lambda_3(u_i - \xi_3)_+ + \\
& \sum_{i=1}^{29} \frac{\frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_1} K\left(\frac{v_1 - v_{1i}}{\phi_1}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_2} K\left(\frac{v_2 - v_{2i}}{\phi_2}\right)} y_i +
\end{aligned}$$

$$\sum_{i=1}^{29} \frac{\frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_3} K\left(\frac{v_3 - v_{3i}}{\phi_3}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right)}{\sum_{i=1}^{29} \frac{1}{\phi_4} K\left(\frac{v_4 - v_{4i}}{\phi_4}\right)} y_i + \varepsilon_i.$$

Syntax R yang digunakan untuk mendapatkan GCV model dengan tiga titik knot disajikan pada Lampiran 6. Lokasi titik-titik knot, bandwidth dan GCV yang diperoleh disajikan pada tabel 4.4.

Tabel 4.4 GCV pada Model dengan Tiga Titik Knot

No	Titik Knot			Bandwidth				GCV
	$\xi_1$	$\xi_2$	$\xi_3$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	
1	9,3142	23,6320	29,3591	0,1181	0,0577	0,0383	0,0076	21,7166
2	12,1778	23,6320	26,4956	0,1065	0,0562	0,0379	0,0076	20,9848
3	15,0413	23,6320	29,3591	0,1142	0,0570	0,0382	0,0076	20,7097
4	17,9049	23,6320	26,4956	0,1016	0,0567	0,0399	0,0076	19,3065
5	20,7685	23,6320	26,4956	0,1097	0,0585	0,0448	0,0076	17,7708
6	23,6320	26,4956	40,8134	0,1019	0,0637	0,0377	0,0036	19,2524
7	<b>26,4956</b>	<b>29,3591</b>	<b>35,0863</b>	<b>0,1411</b>	<b>0,0676</b>	<b>0,0344</b>	<b>0,0207</b>	<b>13,4836</b>
8	26,4956	29,3591	40,8134	0,1357	0,0853	0,0361	0,0228	16,1498
9	29,3591	35,0863	37,9498	0,1064	0,0877	0,0401	0,0072	22,7724
10	32,2227	37,9498	43,6769	0,1065	0,0891	0,0398	0,0092	22,1233

Tabel 4.4 menunjukkan bahwa nilai GCV minimum yang dihasilkan adalah 13,4836. Nilai tersebut diperoleh ketika lokasi titik-titik knot ( $\xi_1, \xi_2, \xi_3$ ) berada pada 26,4956, 29,3591 dan 35,0863 dan bandwidth ( $\phi_1, \phi_2, \phi_3, \phi_4$ ) sama dengan 0,1411, 0,0676, 0,0344 dan 0,0207.

#### 4.5.3.4 GCV Minimum dan Banyaknya Titik Knot Optimum

Perbandingan GCV minimum antara model dengan satu titik knot, model dengan dua titik knot dan model dengan tiga titik knot ditunjukkan oleh tabel 4.5.

Tabel 4.5 Perbandingan nilai GCV Minimum

No	Model	GCV
1	Model dengan Satu Titik Knot	21,2932
2	Model dengan Dua Titik Knot	17,1844
3	<b>Model dengan Tiga Titik Knot</b>	<b>13,4836</b>

Pada Tabel 4.5 terlihat bahwa model regresi yang memiliki GCV minimum adalah model dengan tiga titik knot. Lokasi titik-titik knot dan bandwidth pada GCV minimum tersebut merupakan lokasi titik-titik knot dan bandwidth optimum. Lokasi titik-titik knot dan bandwidth optimum tersebut adalah

$$\begin{aligned}\xi_1 &= 26,4956, & \phi_1 &= 0,1411, \\ \xi_2 &= 29,3591, & \phi_2 &= 0,0676, \\ \xi_3 &= 35,0863, & \phi_3 &= 0,0344, \\ && \phi_4 &= 0,0207.\end{aligned}$$

#### 4.5.4 Estimasi Model Regresi

Model regresi campuran nonparametrik spline dan kernel terbaik diperoleh ketika model menggunakan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum. Berdasarkan pemilihan titik-titik knot, model regresi campuran nonparametrik terbaik diperoleh ketika kurva regresi spline memiliki tiga titik knot. Model berdasarkan tiga titik knot optimum  $\xi_1 = 26,4956$ ,  $\xi_2 = 29,3591$ ,  $\xi_3 = 35,0863$  dan empat bandwidth optimum  $\phi_1 = 0,1411$ ,  $\phi_2 = 0,0676$ ,  $\phi_3 = 0,0344$ ,  $\phi_4 = 0,0207$  adalah

$$y_i = \beta_0 + \beta_1 u_i + \lambda_1(u_i - 26,4956)_+ + \lambda_2(u_i - 29,3591)_+ + \lambda_3(u_i - 35,0863)_+ + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right)} y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0676} K\left(\frac{v_2 - v_{2i}}{0,0676}\right)} y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right)} y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0207} K\left(\frac{v_4 - v_{4i}}{0,0207}\right)} y_i + \varepsilon_i.$$

Hasil estimasi parameter  $\beta_0, \beta_1, \lambda_1, \lambda_2, \lambda_3$  berdasarkan titik-titik knot dan bandwidth tersebut adalah

$$\begin{aligned}\hat{\beta}_0 &= 3,6965, & \hat{\lambda}_1 &= 3,6986, \\ \hat{\beta}_1 &= -0,2832, & \hat{\lambda}_2 &= -5,5289, \\ && \hat{\lambda}_3 &= 2,3139.\end{aligned}$$

Dengan demikian estimasi kurva regresi model menjadi

$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = 3,6965 - 0,2832u_i + 3,6986(u_i - 26,4956)_+ -$$

$$\begin{aligned}
& 5,5289(u_i - 29,3591)_+ + 2,3139(u_i - 35,0863)_+ + \\
& \sum_{i=1}^{29} \frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right) y_i + \sum_{i=1}^{29} \frac{1}{0,0676} K\left(\frac{v_2 - v_{2i}}{0,0676}\right) y_i + \\
& \sum_{i=1}^{29} \frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right) y_i + \sum_{i=1}^{29} \frac{1}{0,0207} K\left(\frac{v_4 - v_{4i}}{0,0207}\right) y_i .
\end{aligned}$$

Dari hasil penghitungan menggunakan *syntax R* pada Lampiran 7 diperoleh nilai  $R^2$  sebesar 0,9202. Nilai  $R^2$  ini menunjukkan bahwa variabel yang digunakan dapat menjelaskan model sebesar 92,02%.

#### 4.5.5 Uji Normalitas Residual

Nilai residual diperoleh dari *syntax R* pada Lampiran 8. Pengujian asumsi kenormalan residual dilakukan dengan menggunakan uji kenormalan Kolmogorov-Smirnov. Dari proses pengujian diperoleh plot pada Gambar 4.2.



Gambar 4.2 *Probability Plot Residual*

Pengujian normalitas pada residual menghasilkan  $p\text{-value} > 0,150$ , lebih besar dari  $\alpha$  (0,05) sehingga disimpulkan bahwa residual model berdistribusi normal.

#### 4.5.6 Interpretasi Model

Model untuk pola data yang mengikuti kernel (gini ratio, rata-rata lama sekolah, tingkat pengangguran terbuka dan laju pertumbuhan ekonomi) tidak dapat diinterpretasi, namun pola data yang mengikuti spline (PDRB perkapita) masih dapat diinterpretasi. Untuk menginterpretasi pola data PDRB perkapita, data PDRB perkapita terlebih dahulu dibagi menjadi 4 bagian, yaitu data yang PDRB perkapitanya kurang dari 26,4956 juta rupiah, di antara 26,4956-29,3591 juta rupiah, di antara 29,3591-35,0863 juta rupiah dan besar dari 35,0863 juta rupiah. Berikut ini adalah model masing-masing kelompok data tersebut dan interpretasi modelnya.

1. Model untuk data PDRB perkapita kurang dari 26,4956 juta rupiah adalah

$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = 3,6965 - 0,2832u_i + \sum_{i=1}^{29} \frac{\frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right)}{\sum_{i=1}^{29} \frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right)} y_i + \\ \sum_{i=1}^{29} \frac{\frac{1}{0,0676} K\left(\frac{v_2 - v_{2i}}{0,0676}\right)}{\sum_{i=1}^{29} \frac{1}{0,0676} K\left(\frac{v_2 - v_{2i}}{0,0676}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right)}{\sum_{i=1}^{29} \frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right)} y_i + \\ \sum_{i=1}^{29} \frac{\frac{1}{0,0207} K\left(\frac{v_4 - v_{4i}}{0,0207}\right)}{\sum_{i=1}^{29} \frac{1}{0,0207} K\left(\frac{v_4 - v_{4i}}{0,0207}\right)} y_i .$$

Dengan mengasumsikan bahwa semua data yang mengikuti kernel konstan, maka ketika PDRB perkapita naik sebesar 1 juta rupiah dan nilainya kurang dari 26,4956 juta rupiah, persentase penduduk miskin akan turun sebesar 0,2832 persen.

2. Model untuk data PDRB perkapita antara 26,4956-29,3591 juta rupiah adalah

$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = 101,6931 + 3,4154u_i + \sum_{i=1}^{29} \frac{\frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right)}{\sum_{i=1}^{29} \frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right)} y_i + \\ \sum_{i=1}^{29} \frac{\frac{1}{0,0676} K\left(\frac{v_2 - v_{2i}}{0,0676}\right)}{\sum_{i=1}^{29} \frac{1}{0,0676} K\left(\frac{v_2 - v_{2i}}{0,0676}\right)} y_i + \sum_{i=1}^{29} \frac{\frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right)}{\sum_{i=1}^{29} \frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right)} y_i +$$

$$\sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0207} K \left( \frac{v_4 - v_{4i}}{0,0207} \right)} y_i.$$

Dengan mengasumsikan bahwa data yang mengikuti kernel konstan, maka ketika PDRB perkapita naik sebesar 1 juta rupiah dan nilainya berada di rentang 26,4956 juta rupiah sampai dengan 29,3591 juta rupiah, persentase penduduk miskin akan naik sebesar 3,4154 persen.

3. Model untuk data PDRB perkapita antara 29,3591-35,0863 juta rupiah adalah

$$\begin{aligned} \hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = & -60,6304 - 2,1135u_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,1411} K \left( \frac{v_1 - v_{1i}}{0,1411} \right)} y_i + \\ & \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0676} K \left( \frac{v_2 - v_{2i}}{0,0676} \right)} y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0344} K \left( \frac{v_3 - v_{3i}}{0,0344} \right)} y_i + \\ & \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0207} K \left( \frac{v_4 - v_{4i}}{0,0207} \right)} y_i. \end{aligned}$$

Dengan mengasumsikan bahwa data yang mengikuti kernel konstan, maka ketika PDRB perkapita naik sebesar 1 juta rupiah dan nilainya berada di rentang 29,3591 juta rupiah dan 35,0863 juta rupiah, persentase penduduk miskin akan turun sebesar 2,1135 persen.

4. Model untuk data PDRB perkapita besar dari 35,0863 juta rupiah adalah

$$\begin{aligned} \hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = & 20,5558 + 0,2004u_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,1411} K \left( \frac{v_1 - v_{1i}}{0,1411} \right)} y_i + \\ & \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0676} K \left( \frac{v_2 - v_{2i}}{0,0676} \right)} y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0344} K \left( \frac{v_3 - v_{3i}}{0,0344} \right)} y_i + \\ & \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0207} K \left( \frac{v_4 - v_{4i}}{0,0207} \right)} y_i. \end{aligned}$$

Dengan mengasumsikan bahwa data yang mengikuti kernel konstan, maka ketika PDRB perkapita naik sebesar 1 juta rupiah dan nilainya lebih besar dari

35,0863 juta rupiah, persentase penduduk miskin akan naik sebesar 0,2004 persen.

Jika dilihat dari interpretasi-interpretasi tersebut, interpretasi model untuk data PDRB perkapita antara 26,4956-29,3591 juta rupiah dan model untuk data PDRB perkapita lebih besar dari 35,0863 juta rupiah tidak sesuai dengan teori, dimana ketika PDRB perkapita mengalami kenaikan, persentase kemiskinan juga mengalami kenaikan. Jika dilihat data PDRB perkapita kabupaten/kotanya, terdapat dua kabupaten/kota yang PDRB perkapitanya berada pada rentang 26,4956-29,3591 juta rupiah, yaitu Kabupaten Supiori dan Kabupaten Mamberamo Tengah. Dua kabupaten ini merupakan dua kabupaten pemekaran yang jumlah penduduknya paling kecil di Provinsi Papua. Jumlah penduduk di Kabupaten Supiori pada tahun 2013 hanya 16.976 jiwa, sedangkan di Kabupaten Mamberamo Tengah hanya 19.776 jiwa. Sektor penyumbang PDRB terbesar di kabupaten-kabupaten tersebut adalah sektor jasa pemerintahan dan bangunan. Hal ini menyebabkan PDRB perkapita di daerah tersebut menjadi tinggi namun tidak menyebabkan kemiskinannya menurun. Selain itu, jika dilihat kabupaten/kota yang PDRB perkapitanya lebih besar dari 35,0863 juta rupiah, terdapat dua kabupaten/kota yang datanya berada di rentang tersebut, yaitu Kabupaten Boven Digoel dan Kota Jayapura. Dimungkinkan ketidaksesuaian teori antara PDRB perkapita dan persentase penduduk miskin disebabkan karena daerah tersebut memiliki PDRB yang tinggi namun pendapatan penduduk tidak merata.

Kemungkinan lain terjadinya ketidaksesuaian teori antara PDRB perkapita dan persentase penduduk miskin di wilayah tersebut adalah adanya dua variabel prediktor di dalam model yang mempunyai sumber data dasar yang sama. Dua variabel prediktor tersebut adalah PDRB perkapita dan laju pertumbuhan ekonomi. Untuk memeriksa adanya pengaruh dua variabel prediktor yang mempunyai sumber data dasar yang sama tersebut, dilakukan percobaan pemodelan dengan mengeluarkan variabel laju pertumbuhan ekonomi ( $v_4$ ). Model baru yang terbentuk setelah laju perumbuhan ekonomi ( $v_4$ ) dikeluarkan adalah

$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = 4,8863 - 0,3749u_i + 4,8796(u_i - 26,4956)_+ - 7,2691(u_i - 29,3591)_+ + 3,0201(u_i - 35,0863)_+$$

$$\sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,1404} K\left(\frac{v_1 - v_{1i}}{0,1404}\right)} y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0687} K\left(\frac{v_2 - v_{2i}}{0,0687}\right)} y_i + \\ \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0347} K\left(\frac{v_3 - v_{3i}}{0,0347}\right)} y_i.$$

Dari model tersebut, diketahui bahwa interpretasi model tidak mengalami perubahan berarti. Oleh karena itu, adanya dua variabel prediktor yang mempunyai sumber data dasar yang sama (PDRB perkapita dan laju pertumbuhan ekonomi) tidak menyebabkan adanya ketidaksesuaian teori antara PDRB perkapita dan persentase penduduk miskin.

Masalah kemiskinan merupakan masalah yang kompleks, dimana kemiskinan tidak hanya disebabkan oleh satu faktor saja. Oleh karena itu, kondisi ketidaksesuaian antara PDRB perkapita dan persentase penduduk miskin ini masih dianggap wajar terjadi. Dimungkinkan masih ada faktor-faktor lain yang mempengaruhi ketidaksesuaian tersebut, namun tidak tercakup ke dalam variabel-variabel yang digunakan dalam model ini. Oleh karena itu, hal ini bisa menjadi temuan untuk melakukan penelitian lebih lanjut mengenai kabupaten-kabupaten yang berada di wilayah terjadinya ketidaksesuaian tersebut.

#### 4.5.7 Simulasi Data untuk Prediksi

Model regresi campuran nonparametrik spline dan kernel ini dapat digunakan untuk skenario kebijakan. Misalkan suatu kabupaten akan menetapkan target PDRB perkapita 9,39 juta rupiah, gini ratio sebesar 0,37 rata-rata lama sekolah 9,21, tingkat penganguruan terbuka 5,31 dan laju pertumbuhan ekonomi 6,83, maka melalui penghitungan menggunakan model regresi campuran nonparametrik spline dan kernel diperoleh persentase penduduk miskin sebesar 26,02 persen.

Jika target-target yang akan ditetapkan tersebut dianggap tidak menekan persentase penduduk miskin secara signifikan, maka bisa dilakukan simulasi perubahan target dan selanjutnya diprediksi kembali persentase penduduk miskinnya. Tabel 4.6 adalah beberapa contoh skenario data untuk mengetahui

persentase penduduk miskin. Untuk melakukan simulasi dengan menggunakan data yang lain, bisa digunakan syntax R pada Lampiran 7.

Tabel 4.6 Skenario Data untuk Prediksi Persentase Penduduk Miskin

No	PDRB Perkapita	Gini Ratio	Rata-rata Lama Sekolah	Tingkat Pengangguran Terbuka	Laju Pertumbuhan Ekonomi	Persentase Penduduk Miskin
1.	5,1500	0,31	4,07	1,11	9,57	35,46
2.	5,8000	0,31	10,44	4,34	5,14	26,14
3.	6,2300	0,34	3,09	9,67	4,05	34,83
4.	9,8300	0,18	9,53	10,44	4,81	29,23
5.	13,3700	0,18	3,96	10,34	12,64	32,57
6.	13,9400	0,44	5,26	4,08	11,31	32,08
7.	14,5100	0,39	4,93	9,49	11,58	30,05
8.	17,2000	0,23	9,51	8,80	8,42	27,63
9.	18,6500	0,15	9,09	10,28	3,94	24,31
10.	20,4900	0,31	9,83	1,34	7,38	31,63
11.	21,4500	0,32	7,28	10,71	11,52	26,65
12.	26,5300	0,40	6,97	9,68	10,48	16,16
13.	29,5300	0,23	5,71	3,21	9,31	37,58
14.	35,2100	0,30	5,19	6,33	3,20	25,40
15.	36,1400	0,21	8,34	1,76	11,19	33,40
16.	37,5800	0,30	2,62	4,12	10,44	19,50
17.	39,7400	0,15	6,31	8,18	11,74	30,82
18.	40,7300	0,16	10,97	8,71	10,20	21,58
19.	42,8500	0,26	9,72	4,69	4,82	28,85
20.	45,5800	0,39	7,23	2,07	12,01	28,59

## BAB 5 KESIMPULAN

### 5.1 Kesimpulan

Model regresi campuran nonparametrik  $y_i = f(u_i, \tilde{v}_i) + \varepsilon_i, i = 1, 2, \dots, n$ ,  $\tilde{v}_i = (v_{1i}, v_{2i}, \dots, v_{mi})^T$  memiliki kurva regresi bersifat aditif  $f(u_i, \tilde{v}_i) = g(u_i) + \sum_{j=1}^m h_j(v_{ji})$ . Komponen kurva regresi  $g(u_i)$  dihampiri dengan kurva regresi spline linier truncated  $g(u_i) = \beta_0 + \beta_1 u_i + \sum_{l=1}^q \lambda_l (u_i - \xi_l)_+$ , sedangkan komponen kurva regresi  $h_j(v_{ji})$  dihampiri dengan kurva regresi kernel Nadaraya-Watson  $\hat{h}_{j\phi_j}(v_{ji}) = n^{-1} \sum_{i=1}^n W_{\phi_j i}(v_j) y_i$ . Error random  $\varepsilon_i$  mengikuti distribusi normal independen dengan mean 0 dan varian  $\sigma^2$ . Estimator kurva regresi campuran spline dan kernel  $f(u_i, \tilde{v}_i)$  adalah

$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) + \sum_{j=1}^m \hat{h}_{j\phi_j}(v_j),$$

dimana

$$\hat{\theta}(\tilde{\xi}, \tilde{\phi}) = \mathbf{B}(\tilde{\xi}, \tilde{\phi})\tilde{y}, \quad \hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi})\tilde{y},$$

$$\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j) = \mathbf{V}(\tilde{\phi})\tilde{y}, \quad \hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v}) = \mathbf{Z}(\tilde{\xi}, \tilde{\phi})\tilde{y}.$$

Estimator-estimator  $\hat{\theta}(\tilde{\xi}, \tilde{\phi})$ ,  $\hat{g}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$ ,  $\sum_{j=1}^m \hat{h}_{j\phi_j}(v_j)$ , dan  $\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u, \tilde{v})$  adalah estimator-estimator yang bias. Namun estimator-estimator tersebut masih merupakan kelas estimator linier dalam observasi.

Model regresi campuran nonparametrik spline dan kernel diterapkan pada data kemiskinan di Provinsi Papua pada tahun 2013 dimana variabel responnya adalah persentase penduduk miskin ( $y$ ), variabel prediktor yang mengikuti kurva regresi spline adalah PDRB perkapita ( $u$ ), dan variabel-variabel prediktor yang mengikuti kurva regresi kernel adalah gini ratio ( $v_1$ ), rata-rata lama sekolah ( $v_2$ ), tingkat pengangguran terbuka ( $v_3$ ) dan laju pertumbuhan ekonomi ( $v_4$ ). Perbandingan antara GCV antara Model dengan satu titik knot, Model dengan dua titik knot dan Model dengan tiga titik knot menunjukkan bahwa model dengan tiga

titik knot  $\xi_1 = 26,4956$ ,  $\xi_2 = 29,3591$ ,  $\xi_3 = 35,0863$  dan empat bandwidth  $\phi_1 = 0,1411$ ,  $\phi_2 = 0,0676$ ,  $\phi_3 = 0,0344$ ,  $\phi_4 = 0,0207$  memberikan GCV yang minimum. Estimasi kurva regresi campuran nonparametrik spline dan kernel yang dihasilkan dari titik-titik knot dan bandwidth tersebut adalah

$$\hat{f}_{\tilde{\phi}, \tilde{\xi}}(u_i, \tilde{v}_i) = 3,6965 - 0,2832u_i + 3,6986(u_i - 26,4956)_+ - 5,5289(u_i - 29,3591)_+ + 2,3139(u_i - 35,0863)_+ + \sum_{i=1}^{29} \frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right) y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,1411} K\left(\frac{v_1 - v_{1i}}{0,1411}\right)} y_i + \sum_{i=1}^{29} \frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right) y_i + \sum_{i=1}^{29} \frac{1}{\sum_{i=1}^{29} \frac{1}{0,0344} K\left(\frac{v_3 - v_{3i}}{0,0344}\right)} y_i.$$

Model ini memberikan  $R^2$  sebesar 92,02 %. Nilai  $R^2$  ini menunjukkan bahwa variabel yang digunakan dapat menjelaskan model sebesar 92,02%. Model regresi campuran nonparametrik spline dan kernel ini dapat digunakan untuk skenario kebijakan.

## 5.2 Saran

Kurva regresi spline yang digunakan dalam penelitian ini adalah kurva regresi spline linier truncated, sedangkan kurva regresi kernel yang digunakan adalah kurva regresi kernel Nadaraya-Watson. Untuk saran penelitian selanjutnya, dapat dilakukan kajian mengenai model regresi campuran nonparametrik spline dan kernel dimana kurva regresi spline yang digunakan adalah spline kuadratik atau kubik, sedangkan kurva regresi kernel yang digunakan adalah linier konstan. Kurva regresi spline dan kurva regresi kernel juga bisa diganti dengan jenis kurva regresi nonparametrik yang lain, misalnya deret fourier, wavelet dan lain-lain sesuai dengan pola datanya.

## DAFTAR PUSTAKA

- Agenor, P. R. (2004). *Unemployment-Poverty Trade-Offs*. Washington DC: The World Bank.
- Agrawal, P. (2008). "Economic Growth and Poverty Reduction: Evidence from Kazakhstan". *Asian Development Review*, Vol. 24, No. 2, hal. 90-115.
- Bilodeau, M. (1992). "Fourier Smoother and Additive Models". *Canadian Journal of Statistics*, Vol. 20, No.3, hal. 257-269.
- Boediono. (1981). *Teori Pertumbuhan Ekonomi*. Yogyakarta: BPFE Yogyakarta.
- BPS. (2014). *Data dan Informasi Kemiskinan Kabupaten/Kota 2013*. Jakarta: Badan Pusat Statistik.
- Budiantara, I. N. (2009). *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Surabaya: ITS Press.
- Budiantara, I. N., & Mulianah. (2007). "Pemilihan Bandwidth Optimum dalam Regresi Semiparametrik Kernel dan Aplikasinya". *SIGMA: Jurnal Sains dan Teknologi*, Vol. 10, No. 2, hal 159-166.
- Budiantara, I. N., Diana, R., Purhadi, & Darmesto, S. (2011). "Relationship Pattern of Poverty and Unemployment in Indonesia with Bayesian Spline Approach". *International Journal of Basic & Applied Sciences IJBAS-IJENS*, Vol. 11, No. 6, hal. 119-127.
- Budiantara, I. N., Ratna, M., Zain, I., & Wibowo, W. (2012). "Modeling 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., Ratna, M., & Zain, I. (2015). "The Combination of Spline and Kernel Estimator for Nonparametric Regression and Its Properties". *Applied Mathematical Sciences*, Vol. 9, No. 122, hal. 6083-6094.
- Cox, D. D., & O'Sullivan, F. (1996). "Penalized Type Estimator for Generalized Nonparametric Regression". *Journal of Multivariate Analysis*, Vol. 56, No. 10, hal. 185-206.
- Ellies, S. (1994). *The Dimension of Poverty*. Kumarian Press.
- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker, Inc.

- Fan, S. (2003). "Public Investment and Poverty Reduction Case Studies from Asia and Implications for Latin America". *Seminario Internacional: Tendencias Y Desafion Del Gato Publiko Para El Desarollo Agricola Y Rural En America Latina Y El Caribe*. Santo Domingo.
- Formby, J. P., Hoover, G. A., & Hoseong, K. (2000). *Economic Growth and Poverty in the United States: Comparisons of Estimates Based Upon Official Poverty Statistics and Sen's Index of Poverty*. Working Paper No. 00-11-01, Univ. of Alabama, Department of Economics, Finance, and Legal Studies.
- Gasser, T., & Muller, H.-G. (1979). *Kernel Estimation of Regression Functions*. Springer Berlin Heidelberg.
- Gay, D. M. (1990). *Usage Summary for Selected Optimization Routines*. Computing Science Technical Report No. 153, Murray Hill: AT&T Bell Laboratories.
- Hardle, W. (1994). *Applied Nonparametric Regression*. Berlin: Humboldt-Universität zu Berlin.
- Hardle, W., Muller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer-Verlag Berlin Heidelberg.
- Iradian, G. (2005). *Inequality, Poverty, and Growth: Cross-Country Evidence*. IMF Working Paper, 1-39.
- Islamiyati, A., & Budiantara, I. N. (2007). "Model Spline dengan Titik-titik Knots dalam Regresi Nonparametrik". *Jurnal INFERENSI*, Vol. 3, hal. 11-21.
- Jalan, J., & Ravallion, M. (2003). "Estimating The Benefit Incidence of An Antipoverty Program". *Journal of Business & Economic Statistics*, Vol. 21, No. 1, hal. 19-30.
- Jung, S. (2009). *Analyzing Poverty in the Southern United States*. Master's Theses, Knoxville: University of Tennessee.
- Kayri, M., & Zirhhoglu. (2009). "Kernel Smoothing Function and Choosing Bandwidth for Nonparametric Regression Methods". *Ozean Journal of Applied Sciences*, 2(1), 49-54.
- Klemela, J. (2014). *Multivariate Nonparametric Regression and Visualization: with R and Applications to Finance*. New Jersey: John Wiley & Sons, Inc.
- Lestari, B., & Budiantara, I. N. (2010). "Spline Estimator of Triple Response Nonparametric Regression Model". *Jurnal Ilmu Dasar*, Vol. 11, hal. 17-22.
- Liang, H. (2006). "Estimation in Partially Linear Models and Numerical Comparisons". *Computational Statistics & Data Analysis*, Vol. 50, No .3, hal. 675-687.

- Lin, Y., & Zhang, H. H. (2006). "Component Selection and Smoothing in Multivariate Nonparametric Regression.". *The Annals of Statistics*, Vol. 34, No. 5, hal. 2272-2297.
- 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 POMITS*, Vol. 2, No. 1, hal. 2337-3520.
- Minoiu, C., & Reddy, S. G. (2008). *Kernel Density Estimation Based on Grouped Data: The Case of Poverty Assessment*. International Monetary Fund.
- Morton, J., & Silverberg, L. (1999). "Fourier Series of Half-range Functions by Smooth Extension". *Applied Mathematical Modelling*, Vol. 33, No. 2, hal. 812-821.
- Muller, C. (1997). *Transient Seasonal and Chronic Poverty of Peasants Evidence from Rwanda*. Oxford: Centre for the Study of African Economies.
- Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwcer Academic Publishers.
- Naja, A. H. (2006). "Pendidikan Berkualitas dan Pembangunan Sumber Daya Manusia: Solusi Utama Masalah Pengangguran dan Kemiskinan di Indonesia". *Jurnal Bisnis dan Ekonomi Politik*, Vol. 7 No. 1, hal. 67-79.
- Nash, J. C., & Varadhan, N. (2011). "Unifying Optimization Algorithms to Aid Software System Users: optimx for R". *Journal of Statistical Software*, Vol. 43, Issue 9.
- Ratnasari, V., Budiantara, I. N., Zain, I., Ratna, M., & Mirah, M. N. (2015). "Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Models (Applications: Data of Poverty in Papua, Indonesia)". *International Journal of Basic & Applied Sciences IJBAS-IJENS*, Vol. 15, No. 04, hal. 9-12.
- Reinsch, C. H. (1967). "Smoothing by Spline Functions". *Numerische Mathematik*, Vol. 10, hal. 77-183.
- Rencher, A. C., & Schaalje, G. B. (2008). *Linears Models in Statistics* (2nd ed.). New Jersey: A John Wiley & Sons.
- Sahdan, G. (2005). "Menanggulangi Kemiskinan di Desa". *Jurnal Ekonomi Rakyat dan Kemiskinan*, Vol. 3, hal. 1-10.
- Sala-i-Martin, X. (2006). "The World Distribution of Income: Falling Poverty and ... Convergence, Period". *The Quarterly Journal of Economics*, Vol. 121, No. 2, hal. 351-397.

- Silverman, B. W. (1985). "Some Aspects of The Spline Smoothing Approach to Non-parametric Regression Curve Fitting". *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 47, No. 1, hal. 1-52.
- Sudiarsa, I. W., Budiantara, I. N., Suhartono, & Purnami, S. W. (2015). "Combined Estimator Fourier Series and Spline Truncated Multivariable Nonparametric Regression". *Applied Mathematical Sciences*, Vol. 9, No. 100, hal. 4997-5010.
- Suharto, E. (2009). *Kemiskinan dan Perlindungan Sosial di Indonesia: Menggagas Model Jaminan Sosial Universal Bidang Kesehatan*. Bandung: Alfabeta.
- Tilak, J. B. (2005). *Post-Elementary Education, Poverty and Development in India*. New Delhi: Working Paper Series - No. 6, Centre of African Studies, University of Edinburgh.
- Tripena, A., & Budiantara, I. N. (2007). "Fourier Estimator in Nonparametric Regression". *International Conference on Natural Sciences and Applied Natural Sciences*. Ahmad Dahlan University, Yogyakarta.
- Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.
- Wand, M. P., & Jones, M. C. (1995). *Kernel Smoothing*. London: Chapman & Hall.
- You, J., & Chen, G. (2007). "Semiparametric Generalized Least Squares Estimation in Partially Linear Regression Models with Correlated Errors". *Journal of Statistical Planning and Inference*, Vol. 137, No. 1, hal. 117-132.

## LAMPIRAN

**Lampiran 1. Data dan Format Data yang Digunakan**

kabupaten/kota	y	u	v1	v2	v3	v4
Merauke	12.33	25.0031	0.466	9.47	4.95	10.24
Jayawijaya	41.81	7.8262	0.367	5.32	0.74	8.14
Jayapura	17.58	24.8528	0.268	9.79	6.43	10.09
Nabire	27.69	18.4047	0.326	7.33	4.58	9.56
Kepulauan Yapen	29.32	11.0291	0.322	6.76	6.14	5.12
Biak Numfor	30.28	15.9910	0.416	9.67	8.81	5.67
Paniai	40.15	3.9398	0.319	6.23	0.00	3.86
Puncak Jaya	39.92	6.6341	0.216	6.13	0.92	3.05
Mimika	20.37	21.6663	0.299	6.94	6.76	8.53
Boven Digoel	23.70	37.9145	0.299	3.65	3.90	5.60
Mappi	30.35	14.4820	0.295	4.37	2.15	11.83
Asmat	33.84	11.2151	0.389	4.44	0.88	6.61
Yahukimo	43.27	3.7336	0.271	2.93	0.22	5.55
Pegunungan Bintang	37.23	13.7145	0.146	2.62	3.11	8.47
Tolikara	38.00	5.1525	0.198	3.48	2.97	4.03
Sarmi	17.72	33.7857	0.328	7.02	4.26	9.30
Keerom	23.23	23.7640	0.283	7.45	5.17	7.66
Waropen	37.27	20.0510	0.221	6.56	5.08	10.16
Supiori	41.50	27.9680	0.280	8.15	11.98	4.73
Mamberamo Raya	34.25	27.7555	0.309	5.21	3.21	8.87
Nduga	39.69	3.5871	0.199	2.82	1.47	12.69
Lanny Jaya	43.79	3.9619	0.181	3.75	0.00	7.36
Mamberamo Tengah	39.59	7.4219	0.234	2.94	0.00	11.66
Yalimo	40.33	5.1929	0.259	2.85	0.15	12.21
Puncak	41.96	9.2934	0.183	2.86	2.04	12.61
Dogiyai	32.25	8.5494	0.192	4.17	3.88	6.67
Intan Jaya	42.03	7.5741	0.225	2.30	0.75	7.58
Deiyai	47.52	5.1668	0.319	2.97	0.00	4.87
Kota Jayapura	16.19	46.5405	0.375	11.07	9.84	12.28

<b>Keterangan:</b> y = persentase penduduk miskin u = PDRB perkapita v1 = gini ratio v2 = rata-rata lama sekolah v3 = tingkat pengangguran terbuka v4 = laju pertumbuhan ekonomi
--

## Lampiran 2. Fungsi nlminb pada Program R

```

function (start, objective, gradient = NULL, hessian = NULL,
..., scale = 1, control = list(), lower = -Inf, upper = Inf)
{
  par <- setNames(as.double(start), names(start))
  n <- length(par)
  iv <- integer(78 + 3 * n)
  v <- double(130 + (n * (n + 27))/2)
  .Call(C_port_ivset, 2, iv, v)
  if (length(control)) {
    nms <- names(control)
    if (!is.list(control) || is.null(nms))
      stop("'control' argument must be a named list")
    pos <- pmatch(nms, names(port_cpos))
    if (any(nap <- is.na(pos))) {
      warning(sprintfgettext(length(nap), "unrecognized control
element named %s ignored",
"unrecognized control elements named %s ignored"),
       paste(sQuote(nms[nap]), collapse = ", ")), domain = NA)
      pos <- pos[!nap]
      control <- control[!nap]
    }
    ivpars <- pos <= 4
    vpars <- !ivpars
    if (any(ivpars))
      iv[port_cpos[pos[ivpars]]] <- as.integer(unlist(control[ivpars]))
    if (any(vpars))
      v[port_cpos[pos[vpars]]] <- as.double(unlist(control[vpars]))
  }
  obj <- quote(objective(.par, ...))
  rho <- new.env(parent = environment())
  assign(".par", par, envir = rho)
  grad <- Hess <- low <- upp <- NULL
  if (!is.null(gradient)) {
    grad <- quote(gradient(.par, ...))
    if (!is.null(hessian)) {
      if (is.logical(hessian))
        stop("logical 'hessian' argument not allowed. See
documentation.")
      Hess <- quote(hessian(.par, ...))
    }
  }
  if (any(lower != -Inf) || any(upper != Inf)) {
    low <- rep_len(as.double(lower), length(par))
    upp <- rep_len(as.double(upper), length(par))
  }
  else low <- upp <- numeric()
  .Call(C_port_nlminb, obj, grad, Hess, rho, low, upp, d =
rep_len(as.double(scale),
length(par)), iv, v)
  iv1 <- iv[1L]
  list(par = get(".par", envir = rho), objective = v[10L],
convergence = (if (iv1 %%in% 3L:6L) 0L else 1L), iterations = iv[31L],
evaluations = c(`function` = iv[6L], gradient = iv[30L]),
message = if (19 <= iv1 && iv1 <= 43) {
  if (any(B <- iv1 == port_cpos)) sprintf("'control' component
'%s' = %g, is out of range",
names(port_cpos)[B], v[iv1]) else sprintf("V[IV[1]] = "
)
}
}

```

```
V[%d] = %g is out of range (see PORT docu.)",
    iv1, v[iv1])
} else port_msg(iv1)
}
```

### Lampiran 3. Prosedur Pemanggilan Fungsi nlminb

nlminb {stats}		R Documentation		
Optimization using PORT routines				
<b>Description</b>				
Unconstrained and box-constrained optimization using PORT routines. For historical compatibility.				
<b>Usage</b>				
<pre>nlminb(start, objective, gradient = NULL, hessian = NULL, ...,        scale = 1, control = list(), lower = -Inf, upper = Inf)</pre>				
<b>Arguments</b>				
start	numeric vector, initial values for the parameters to be optimized.			
objective	Function to be minimized. Must return a scalar value. The first argument to objective is the vector of parameters to be optimized, whose initial values are supplied through start. Further arguments (fixed during the course of the optimization) to objective may be specified as well (see ...).			
gradient	Optional function that takes the same arguments as objective and evaluates the gradient of objective at its first argument. Must return a vector as long as start.			
hessian	Optional function that takes the same arguments as objective and evaluates the hessian of objective at its first argument. Must return a square matrix of order length(start). Only the lower triangle is used.			
...	Further arguments to be supplied to objective.			
scale	See PORT documentation (or leave alone).			
control	A list of control parameters. See below for details.			
lower, upper	vectors of lower and upper bounds, replicated to be as long as start. If unspecified, all parameters are assumed to be unconstrained.			
<b>Details</b>				
Any names of start are passed on to objective and where applicable, gradient and hessian. The parameter vector will be coerced to double.				
The PORT documentation is at <a href="http://netlib.bell-labs.com/cm/cs/cstr/153.pdf">http://netlib.bell-labs.com/cm/cs/cstr/153.pdf</a> .				
The parameter vector passed to objective, gradient and hessian had special semantics prior to R 3.1.0 and was shared between calls. The functions should not copy it.				
If any of the functions returns NA or NaN the internal code could infinite-loop in R prior to 2.15.2: this is now an error for the gradient and Hessian, and such values for function evaluation are replaced by +Inf with a warning.				
<b>Value</b>				
A list with components:				
par	The best set of parameters found.			
objective	The value of objective corresponding to par.			
convergence	An integer code. 0 indicates successful convergence.			
message	A character string giving any additional information returned by the optimizer, or NULL. For details, see PORT documentation.			
iterations	Number of iterations performed.			

**evaluations** Number of objective function and gradient function evaluations

**Control parameters**

Possible names in the control list and their default values are:

- eval.max**  
Maximum number of evaluations of the objective function allowed.  
Defaults to 200.
- iter.max**  
Maximum number of iterations allowed. Defaults to 150.
- trace**  
The value of the objective function and the parameters is printed every trace'th iteration. Defaults to 0 which indicates no trace information is to be printed.
- abs.tol**  
Absolute tolerance. Defaults to 0 so the absolute convergence test is not used. If the objective function is known to be non-negative, the previous default of 1e-20 would be more appropriate.
- rel.tol**  
Relative tolerance. Defaults to 1e-10.
- x.tol**  
X tolerance. Defaults to 1.5e-8.
- xf.tol**  
false convergence tolerance. Defaults to 2.2e-14.
- step.min, step.max**  
Minimum and maximum step size. Both default to 1..
- sing.tol**  
singular convergence tolerance; defaults to rel.tol.
- scale.init**  
...
- diff.g**  
an estimated bound on the relative error in the objective function value.

#### Lampiran 4. Syntax R GCV Model dengan Satu Titik Knot

```

library(pracma)
kemiskinan <- read.csv("datakemiskinan.csv", header=TRUE)
n <- dim(kemiskinan)[1]
y <- kemiskinan[,c(2)]
u <- kemiskinan[,c(3)]
v <- kemiskinan[,c(4,5,6,7)]
m <- ncol(v)
m1.nn <- matrix(1, nrow=n, ncol=n)
m1.n1 <- matrix(1, nrow=n, ncol=1)
mi.nn <- diag(1, nrow=n, ncol=n)
int <- 561
XI <- seq(min(u), max(u), (max(u)-min(u))/(int-1))
PHI <- matrix(1, nrow=int, ncol=4)
GCV <- matrix(1, nrow=int, ncol=1)
MSE <- matrix(1, nrow=int, ncol=1)
for(i in 1:int)
{
  xi <- XI[i]
  GCVmin <- function(h)
  {
    phi <- h[c(1,2,3,4)]
    sum.V.phi <- 0
    for(j in 1:m)
    {
      v.diag <- diag(v[,j])
      V <- m1.nn%*%v.diag
      Z <- (t(V)-V)/phi[j]
      K <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
      K.Z <- (1/phi[j])*K
      W.penyebut <- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
      V.phi <- 1/n*K.Z/W.penyebut
      sum.V.phi <- sum.V.phi+V.phi
    }
    V.PHI <- sum.V.phi/m
    G.XI <- cbind(1, u, pmax(0, u-xi))
    S.XI.PHI <- G.XI%*%pinv(t(G.XI)%*%G.XI)%*%t(G.XI)%*%(mi.nn-V.PHI)
    Z.XI.PHI <- S.XI.PHI+V.PHI
    f.cap <- Z.XI.PHI%*%y
    SSE <- sum((y-f.cap)^2)
    MSE <- SSE/n
    GCV <- MSE/(n^-1*sum(diag(mi.nn-Z.XI.PHI)^2))
  }
  optimasi <- nlm(function(phi) MSE, start=c(0.05, 4), lower=0)
  PHI[i,] <- optimasi$par
  GCV[i] <- optimasi$objective
  phi <- optimasi$par[c(1,2,3,4)]
  sum.V.phi <- 0
  for(j in 1:m)
  {
    v.diag <- diag(v[,j])
    V <- m1.nn%*%v.diag
    Z <- (t(V)-V)/phi[j]
    K <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
    K.Z <- (1/phi[j])*K
    W.penyebut <- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
    V.phi <- 1/n*K.Z/W.penyebut
  }
}

```

```

    sum.V.phi      <- sum.V.phi+V.phi
}
V.PHI          <- sum.V.phi/m
G.XI           <- cbind(1,u,pmax(0,u-xi))
S.XI.PHI       <- G.XI%*%pinv(t(G.XI)%*%G.XI)%*%t(G.XI)%*%(mi.nn-V.PHI)
Z.XI.PHI       <- S.XI.PHI+V.PHI
f.cap          <- Z.XI.PHI%*%y
SSE            <- sum((y-f.cap)^2)
MSE[i]         <- SSE/n
}

GCV.XIPHI      <- cbind(XI,PHI,MSE,GCV)
colnames(GCV.XIPHI) <- c("xi","phi1","phi2","phi3","phi4","MSE","GCV")
GCV.XIPHI      <- GCV.XIPHI[order(GCV,MSE),]
write.csv(GCV.XIPHI,file="GCV_1_knot.csv")

```

## Lampiran 5. Syntax R GCV Model dengan Dua Titik Knot

```

library(pracma)
kemiskinan <- read.csv("datakemiskinan.csv", header=TRUE)
n <- dim(kemiskinan)[1]
y <- kemiskinan[,c(2)]
u <- kemiskinan[,c(3)]
v <- kemiskinan[,c(4,5,6,7)]
m <- ncol(v)
m1.nn <- matrix(1, nrow=n, ncol=n)
m1.n1 <- matrix(1, nrow=n, ncol=1)
mi.nn <- diag(1, nrow=n, ncol=n)
int <- 34
xi <- seq(min(u), max(u), (max(u)-min(u))/(int-1))
n.komb <- 0
for(r in 1:(int-1))
{
  n.komb <- n.komb+r
}
XI <- matrix(1, nrow=n.komb, ncol=2)
PHI <- matrix(1, nrow=n.komb, ncol=4)
GCV <- matrix(1, nrow=n.komb, ncol=1)
MSE <- matrix(1, nrow=n.komb, ncol=1)
r <- 1
for(i in 1:(int-1))
{
  for(k in (i+1):int)
  {
    GCVmin <- function(h)
    {
      phi <- h[c(1,2,3,4)]
      <- 0
      sum.V.phi <- 0
      for(j in 1:m)
      {
        v.diag <- diag(v[,j])
        V <- m1.nn%*%v.diag
        Z <- (t(V)-V)/phi[j]
        K <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
        K.Z <- (1/phi[j])*K
        W.penyebut <- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
        <- 1/n*K.Z/W.penyebut
        V.phi <- sum.V.phi+V.phi
      }
      V.PHI <- sum.V.phi/m
      G.XI <- cbind(1,u,pmax(0,u-xi[i]),pmax(0,u-xi[k]))
      S.XI.PHI <- G.XI%*%pinv(t(G.XI)%*%G.XI)%*%t(G.XI)%*%
        (mi.nn-V.PHI)
      Z.XI.PHI <- S.XI.PHI+V.PHI
      f.cap <- Z.XI.PHI%*%y
      SSE <- sum((y-f.cap)^2)
      MSE <- SSE/n
      GCV <- MSE/(n^-1*sum(diag(mi.nn-Z.XI.PHI)^2))
    }
    optimasi <- nlmib(c(rep(0.05,4)),GCVmin,lower=0)
    XI[r,] <- c(xi[i],xi[k])
    PHI[r,] <- optimasi$par
    GCV[r] <- optimasi$objective
    phi <- optimasi$par[c(1,2,3,4)]
  }
}

```

```

sum.V.phi      <- 0
for(j in 1:m) {
  v.diag
  V
  Z
  K
  K.Z
  W.penyebut
  V.phi
  sum.V.phi
}
V.PHI
G.XI
S.XI.PHI
Z.XI.PHI
f.cap
SSE
MSE[r]
r
}
GCV.XIPHI
colnames(GCV.XIPHI)
GCV.XIPHI
write.csv(GCV.XIPHI,file="GCV_2_knot.csv")
<- diag(v[,j])
<- m1.nn%*%v.diag
<- (t(V)-V)/phi[j]
<- 1/sqrt(2*pi)*exp(-1/2*Z^2)
<- (1/phi[j])*K
<- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
<- 1/n*K.Z/W.penyebut
<- sum.V.phi+V.phi
<- sum.V.phi/m
<- cbind(1,u,pmax(0,u-xi[i]),pmax(0,u-xi[k]))
<- G.XI%*%pinv(t(G.XI)%*%G.XI)%*%t(G.XI)%*%
(m.i.nn-V.PHI)
<- S.XI.PHI+V.PHI
<- Z.XI.PHI%*%y
<- sum((y-f.cap)^2)
<- SSE/n
<- r+1
<- cbind(XI,PHI,MSE,GCV)
<- c("x11","x12","phi1","phi2","phi3","phi4",
"mse","gcv")
<- GCV.XIPHI[order(GCV,MSE),]

```

## Lampiran 6. Syntax R GCV Model dengan Tiga Titik Knot

```

library(pracma)
kemiskinan <- read.csv("datakemiskinan.csv", header=TRUE)
n <- dim(kemiskinan)[1]
y <- kemiskinan[,c(2)]
u <- kemiskinan[,c(3)]
v <- kemiskinan[,c(4,5,6,7)]
m <- ncol(v)
m1.nn <- matrix(1, nrow=n, ncol=n)
m1.n1 <- matrix(1, nrow=n, ncol=1)
mi.nn <- diag(1, nrow=n, ncol=n)
int <- 16
xi <- seq(min(u), max(u), (max(u)-min(u))/(int-1))
n.komb2 <- 0
n.komb3 <- 0
for(i in 1:(int-2))
{
  n.komb2 <- n.komb2+i
  for(j in (i+1):(int-1))
  {
    n.komb3 <- n.komb3+i
  }
}
XI <- matrix(1, nrow=n.komb3, ncol=3)
PHI <- matrix(1, nrow=n.komb3, ncol=4)
GCV <- matrix(1, nrow=n.komb3, ncol=1)
MSE <- matrix(1, nrow=n.komb3, ncol=1)
r <- 1
for(i in 1:(int-2))
{
  for(k in (i+1):(int-1))
  {
    for(l in (k+1):int)
    {
      GCVmin <- function(h)
      {
        phi <- h[c(1,2,3,4)]
        sum.V.phi <- 0
        for(j in 1:m)
        {
          v.diag <- diag(v[,j])
          V <- m1.nn%*%v.diag
          Z <- (t(V)-V)/phi[j]
          K <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
          K.Z <- (1/phi[j])*K
          W.penyebut <- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
          V.phi <- 1/n*K.Z/W.penyebut
          sum.V.phi <- sum.V.phi+V.phi
        }
        V.PHI <- sum.V.phi/m
        G.XI <- cbind(1, u, pmax(0, u-xi[i]), pmax(0, u-xi[k]),
                      pmax(0, u-xi[1]))
        S.XI.PHI <- G.XI%*%pinv(t(G.XI)%*%G.XI)%*%t(G.XI)%*%
          (mi.nn-V.PHI)
        Z.XI.PHI <- S.XI.PHI+V.PHI
        f.cap <- Z.XI.PHI%*%y
        SSE <- sum((y-f.cap)^2)
      }
    }
  }
}

```

```

MSE
GCV
}
optimasi
XI[r,]
PHI[r,]
GCV[r]
phi
sum.V.phi
for(j in 1:m)
{
  v.diag
  V
  Z
  K
  K.Z
  W.penyebut
  V.phi
  sum.V.phi
}
V.PHI
G.XI
S.XI.PHI
Z.XI.PHI
f.cap
SSE
MSE[r]
r
}
}
GCV.XIPHI
colnames(GCV.XIPHI)
GCV.XIPHI
write.csv(GCV.XIPHI,file="GCV_3_knot.csv")

```

<- SSE/n  
<- MSE/(n^-1\*sum(diag(mi.nn-Z.XI.PHI)^2))  
<- nlminb(c(rep(0.05,4)),GCVmin,lower=0)  
<- c(xi[i],xi[k],xi[l])  
<- optimasi\$par  
<- optimasi\$objective  
<- optimasi\$par[c(1,2,3,4)]  
<- 0  
<- diag(v[,j])  
<- m1.nn%\*%v.diag  
<- (t(V)-V)/phi[j]  
<- 1/sqrt(2\*pi)\*exp(-1/2\*Z^2)  
<- (1/phi[j])\*K  
<- diag(c(1/n\*K.Z%\*%m1.n1))%\*%m1.nn  
<- 1/n\*K.Z/W.penyebut  
<- sum.V.phi+V.phi  
<- sum.V.phi/m  
<- cbind(1,u,pmax(0,u-xi[i]),  
 pmax(0,u-xi[k]),pmax(0,u-xi[l]))  
<- G.XI%\*%pinv(t(G.XI)%\*%G.XI)%\*%t(G.XI)%\*%  
 (mi.nn-V.PHI)  
<- S.XI.PHI+V.PHI  
<- Z.XI.PHI%\*%y  
<- sum(y-f.cap)^2  
<- SSE/n  
<- r+1  
<- cbind(XI,PHI,MSE,GCV)  
<- c("x11","x12","x13","phi1","phi2","phi3",  
 "phi4","MSE","GCV")  
<- GCV.XIPHI[order(GCV),]

## Lampiran 7. Syntax R Estimasi Parameter Model dengan Tiga Titik Knot

```

est.par <- function(xi1,xi2,xi3,phi1,phi2,phi3,phi4)
{
  h <- c(xi1,xi2,xi3,phi1,phi2,phi3,phi4)
  kemiskinan <- read.csv("datakemiskinan.csv",header=TRUE)
  n <- dim(kemiskinan)[1]
  y <- kemiskinan[,c(2)]
  u <- kemiskinan[,c(3)]
  v <- kemiskinan[,c(4,5,6,7)]
  m <- ncol(v)
  m1.nn <- matrix(1,nrow=n,ncol=n)
  m1.n1 <- matrix(1,nrow=n)
  mi.nn <- diag(1,nrow=n,ncol=n)
  xi <- h[c(1,2,3)]
  phi <- h[c(4,5,6,7)]

  sum.V.phi <- 0
  for (j in 1:m)
  {
    v.diag <- diag(v[,j])
    V <- m1.nn%*%v.diag
    Z <- (t(V)-V)/phi[j]
    K <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
    K.Z <- (1/phi[j])*K
    W.penyebut <- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
    V.phi <- 1/n*K.Z/W.penyebut
    sum.V.phi <- sum.V.phi+V.phi
  }
  V.PHI <- sum.V.phi/m
  G.XI <- cbind(1,u,pmax(0,u-xi[1]),pmax(0,u-xi[2]),
                 pmax(0,u-xi[3]))
  S.XI.PHI <- G.XI%*%solve(t(G.XI)%*%G.XI)%*%t(G.XI)%*%(mi.nn-V.PHI)
  Z.XI.PHI <- S.XI.PHI+V.PHI
  f.cap <- Z.XI.PHI%*%y
  SSR <- sum((f.cap-mean(y))^2)
  SST <- sum((y-mean(y))^2)
  MSE <- sum((y-f.cap)^2)/n
  R2 <- SSR/SST
  teta <- solve(t(G.XI)%*%G.XI)%*%t(G.XI)%*%(mi.nn-V.PHI)%*%y
  parameter <- cbind(t(teta),R2)
  colnames(parameter) <- c("beta0","beta1","lambda1","lambda2",
                            "lambda3","R-Squared")
  write.csv(parameter,file="parameter_splinekernel.csv")
  parameter
}

```

## Lampiran 8. Syntax R untuk Menghitung Estimasi $y_i$ dan Residual

```

est.fcap <- function(xi1,xi2,xi3,phi1,phi2,phi3,phi4)
{
  h <- c(xi1,xi2,xi3,phi1,phi2,phi3,phi4)
  kemiskinan <- read.csv("datakemiskinan.csv",header=TRUE)
  n <- dim(kemiskinan)[1]
  y <- kemiskinan[,c(2)]
  u <- kemiskinan[,c(3)]
  v <- kemiskinan[,c(4,5,6,7)]
  m <- ncol(v)
  m1.nn <- matrix(1,nrow=n,ncol=n)
  m1.n1 <- matrix(1,nrow=n)
  mi.nn <- diag(1,nrow=n,ncol=n)
  xi <- h[c(1,2,3)]
  phi <- h[c(4,5,6,7)]

  sum.V.phi <- 0
  for (j in 1:m)
  {
    v.diag <- diag(v[,j])
    V <- m1.nn%*%v.diag
    Z <- (t(V)-V)/phi[j]
    K <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
    K.Z <- (1/phi[j])*K
    W.penyebut <- diag(c(1/n*K.Z%*%m1.n1))%*%m1.nn
    V.phi <- 1/n*K.Z/W.penyebut
    sum.V.phi <- sum.V.phi+V.phi
  }
  V.PHI <- sum.V.phi/m
  G.XI <- cbind(1,u,pmax(0,u-xi[1]),pmax(0,u-xi[2]),
                pmax(0,u-xi[3]))
  S.XI.PHI <- G.XI%*%solve(t(G.XI)%*%G.XI)%*%t(G.XI)%*%(mi.nn-V.PHI)
  Z.XI.PHI <- S.XI.PHI+V.PHI
  f.cap <- Z.XI.PHI%*%y
  SSR <- sum((f.cap-mean(y))^2)
  SST <- sum((y-mean(y))^2)
  MSE <- sum((y-f.cap)^2)/n
  R2 <- SSR/SST
  teta <- solve(t(G.XI)%*%G.XI)%*%t(G.XI)%*%(mi.nn-V.PHI)%*%y
  res <- y-f.cap
  y_fcap_res <- cbind(y,f.cap,res)
  colnames(y_fcap_res) <- c("y","f.cap","residual")
  y_fcap_res
}

```

## Lampiran 9. Syntax R Prediksi Persentase Penduduk Miskin dari Model

```

est.y      <- function(u.est,v1.est,v2.est,v3.est,v4.est)
{
  kemiskinan <- read.csv("datakemiskinan.csv", header=TRUE)
  n           <- dim(kemiskinan)[1]
  y           <- kemiskinan[,c(2)]
  u           <- kemiskinan[,c(3)]
  v           <- kemiskinan[,c(4,5,6,7)]
  m           <- ncol(v)

  u.est       <- u.est
  v.est       <- c(v1.est,v2.est,v3.est,v4.est)

  xi1        <- 26.49558
  xi2        <- 29.35914
  xi3        <- 35.08626
  phi1       <- 0.141065463
  phi2       <- 0.067635127
  phi3       <- 0.034447138
  phi4       <- 0.020748933

  beta.0     <- 3.696494683
  beta.1     <- -0.283172826
  lambda.1   <- 3.698561554
  lambda.2   <- -5.528862823
  lambda.3   <- 2.313874912

  xi         <- c(xi1,xi2,xi3)
  est.spline <- beta.0+(beta.1*u.est)+ 
    (lambda.1*pmax(0,u.est-xi[1]))+
    (lambda.2*pmax(0,u.est-xi[2]))+
    (lambda.3*pmax(0,u.est-xi[3]))

  phi        <- c(phi1,phi2,phi3,phi4)
  sum.V.phi <- 0
  for (j in 1:m)
  {
    x.j       <- matrix(v.est[j], ncol=n)
    Z         <- (x.j-t(v[,j]))/phi[j]
    K         <- 1/sqrt(2*pi)*exp(-1/2*Z^2)
    K.Z       <- (1/phi[j])*K
    W.penyebut <- 1/n*sum(K.Z)
    V.phi     <- 1/n*K.Z/W.penyebut
    sum.V.phi <- sum.V.phi+V.phi
  }
  V.PHI      <- sum.V.phi/m
  est.kernel <- V.PHI%*%y
  f.cap      <- est.spline+est.kernel
  colnames(f.cap) <- c("f.cap")
  f.cap
}

```

## BIOGRAFI PENULIS



Penulis bernama singkat Rory, dilahirkan di Kapau, Kabupaten Agam, Provinsi Sumatera Barat pada tanggal 15 Oktober 1984. Penulis merupakan anak kedua dari dua bersaudara dari pasangan suami-istri Bapak Amziur dan Ibu Erneli. Penulis menyelesaikan pendidikan formal sekolah dasar di SD 01 Padang Panjang Timur pada tahun 1996, kemudian menyelesaikan pendidikan sekolah menengah tingkat pertama di SMP 1 Padang Panjang. Selanjutnya menyelesaikan pendidikan sekolah menengah tingkat atas di SMA 2 Padang Panjang pada tahun 2002. Pada tahun 2002 tersebut, penulis melanjutkan pendidikan ke jenjang perguruan tinggi D-IV di Sekolah Tinggi Ilmu Statistik (STIS) Jakarta yang merupakan perguruan tinggi kedinasan di bawah naungan Badan Pusat Statistik (BPS). Jurusan peminatan yang diambil adalah Statistik Sosial dan Kependudukan. Setelah menyelesaikan pendidikan di STIS pada tahun 2006, penulis bertugas sebagai Pegawai Negeri Sipil (PNS) di BPS Kabupaten Tanjung Jabung Barat, Provinsi Jambi hingga Juli 2014. Pada Agustus 2014, penulis mendapatkan kesempatan untuk melanjutkan pendidikan pada Program Magister Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) di Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Penulis dapat dihubungi di alamat email [rory@bps.go.id](mailto:rory@bps.go.id).

Penulis

Surabaya, Januari 2016