

**MODEL DATA KEMISKINAN DI PROVINSI PAPUA DENGAN REGRESI  
SEMIPARAMETRIK MELALUI *PENALIZED SPLINE* TERBOBOTI**



**NURUL KURUNULBAHRIAH ALIYAH  
H051201066**



**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

**MODEL DATA KEMISKINAN DI PROVINSI PAPUA DENGAN REGRESI  
SEMIPARAMETRIK MELALUI *PENALIZED SPLINE* TERBOBOTI**

**NURUL KURUNULBAHRIAH ALIYAH  
H051201066**



**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

**MODEL DATA KEMISKINAN DI PROVINSI PAPUA DENGAN REGRESI  
SEMIPARAMETRIK MELALUI *PENALIZED SPLINE* TERBOBOTI**

NURUL KURUNULBAHRIAH ALIYAH  
H051201066



Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Program Studi Statistika

pada

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2024**

# SKRIPSI

## MODEL DATA KEMISKINAN DI PROVINSI PAPUA DENGAN REGRESI SEMIPARAMETRIK MELALUI *PENALIZED SPLINE* TERBOBOTI

**NURUL KURUNULBAHRIAH ALIYAH**

**H051201066**

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada 20  
November 2024 dan dinyatakan telah memenuhi syarat kelulusan  
pada

Program Studi Statistika  
Departemen Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin  
Makassar

Mengesahkan:  
Pembimbing Tugas Akhir,



Dr. Anna Islamiyati, S.Si., M. Si.  
NIP. 19770808 200501 2 002

Mengetahui:  
Ketua Program Studi,



Dr. Anna Islamiyati, S.Si., M. Si.  
NIP. 19770808 200501 2 002

## PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Model Data Kemiskinan di Provinsi Papua dengan Regresi Semiparametrik melalui Penalized Spline Terboboti" adalah benar karya saya dengan arahan dari pembimbing Dr. Anna Islamiyati, S.Si., M. Si. sebagai Pembimbing Utama. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas pembuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin

Makassar, 20 November 2024



10000  
SEPULUH RIBU RUPIAH  
TBL  
30  
METER  
TEMPEL  
NIRKOL  
NIRKOL  
NIRKOL

NIRKOL Aliyah  
NIM H051201066

## UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "**Model Data Kemiskinan di Provinsi Papua dengan Regresi Semiparametrik melalui *Penalized Spline* Terboboti**". Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada **Ibu Dr. Anna Islamiyati, S.Si., M. Si.** selaku Pembimbing Utama yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya, memberikan motivasi, dorongan semangat, dan bimbingan kepada penulis selama proses penulisan tugas akhir ini. Terima kasih kepada **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.** dan **Ibu Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.**, selaku Tim Penguji yang senantiasa memberikan saran dan kritikan yang membangun dalam penyempurnaan penulisan tugas akhir ini. Terima kasih kepada **Pimpinan Universitas Hasanuddin, Departemen Statistika, Jajaran Dosen,** dan **Staf Departemen Statistika** yang telah memfasilitasi, memberikan ilmu bermanfaat, dan membantu penulis selama menempuh studi. Tidak lupa pula penulis menyampaikan ucapan terima kasih kepada **Yayasan Hadji Kalla** karena telah memberikan dukungan finansial untuk pendidikan sarjana penulis.

Penghargaan dan ucapan terima kasih yang tulus juga penulis ucapkan kepada kedua orang tua penulis yang tercinta, Ayahanda **Margunawan** dan Ibunda **Rukmini Mudzakir** yang telah memberikan kasih sayang, dukungan, semangat, dan doa yang senantiasa mengiringi langkah penulis. Terima kasih kepada keluarga besar **Ello** dan **Mudzakir** terkhusus saudara-saudara penulis **Muh. Mashury Attamimi** dan **Nurul Faiqah Rezky** yang senantiasa mewarnai hari-hari dan menemani penulis. Terima kasih kepada **Ilmi**, sepupu sekaligus sahabat penulis yang senantiasa menjadi *support system* terbaik bagi penulis. Terima kasih kepada sahabat SMA penulis, **Dea, Afnanin, Insani, Magfira,** dan **Yola** yang telah memberi dukungan dan doa selama pengerjaan tugas akhir ini. Terima kasih kepada sahabat tercinta penulis di Statistika 2020, **Rifdah, Aisyah, Ayu, Dania, Isra, Nurfa, Tiwi, Ira, Dwi,** dan **Aviatul** yang telah menjadi sahabat seperjuangan penulis dengan segala pengalaman berharga selama menempuh studi. Terima kasih kepada keluarga besar **Himastat FMIPA Unhas** khususnya teman-teman **POIS20N** atas kebersamaannya selama ini. Terima kasih kepada teman-teman **KKNT Posko 2 Desa Mappakalombo 110, Rezky, Riri, Ersya,** dan **Chinta** yang telah memberikan dukungan dan semangat kepada penulis. Terima kasih kepada keluarga **Kalla Youth Changemakers 2023** khususnya **Kak Ama, Kelompok 4,** dan **Kelompok 1 Pengabdian Masyarakat Desa Suruang** yang telah memberikan dukungan dan pengalaman yang berharga bagi penulis.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Penulis,



Nurul Kurunulbahriah Aliyah

## ABSTRAK

Nurul Kurunulbahriah Aliyah. **Model Data Kemiskinan di Provinsi Papua dengan Regresi Semiparametrik melalui *Penalized Spline* Terboboti** (dibimbing oleh Anna Islamiyati).

**Latar belakang.** Dalam beberapa kasus, terdapat pola hubungan antara variabel respon dengan satu atau beberapa variabel prediktor diketahui dan sebagian lainnya tidak diketahui. Situasi tersebut dapat dianalisis dengan pendekatan regresi semiparametrik. Salah satu estimator yang dapat digunakan adalah *penalized spline* dengan *Weighted Least Square* (WLS). Penelitian ini diterapkan pada data kemiskinan di Provinsi Papua, dengan variabel prediktor berpola parametrik terjadi pada rata-rata lama sekolah dan pengeluaran perkapita, serta variabel prediktor berpola nonparametrik terjadi pada umur harapan hidup. **Tujuan.** Penelitian ini bertujuan untuk memperoleh estimasi parameter dan membentuk model data kemiskinan di Provinsi Papua dengan regresi semiparametrik melalui *penalized spline* terboboti. **Metode.** Penelitian ini terdiri dari dua tahap, yaitu estimasi parameter regresi semiparametrik melalui *penalized spline* terboboti dan model data kemiskinan di Provinsi Papua dengan regresi semiparametrik melalui *penalized spline* terboboti. **Hasil.** Estimasi parameter regresi semiparametrik melalui *penalized spline* terboboti pada persentase kemiskinan di Provinsi Papua tahun 2016 – 2021 yang diperoleh adalah  $\hat{y} = X\hat{\beta} + Z\hat{\theta}$ . Model dengan nilai GCV minimum terletak pada model dengan orde linear, parameter penghalus sebesar 0,3; dan satu titik knot sebesar 70,52. Model yang diperoleh menunjukkan bahwa apabila umur harapan hidup di bawah 70,52 tahun cenderung mengalami peningkatan persentase penduduk miskin sebesar 0,480%, tetapi setelah melewati 70,52 tahun, persentase penduduk miskin justru menurun secara signifikan. **Kesimpulan.** Dengan demikian, dapat disimpulkan bahwa model regresi semiparametrik melalui *penalized spline* terboboti yang lebih baik terletak pada model yang menggunakan satu titik knot dengan nilai GCV sebesar 4,517.

**Kata kunci:** Regresi Semiparametrik, *Penalized Spline*, *Weighted Least Square*, Kemiskinan

## ABSTRACT

Nurul Kurunulbahriah Aliyah. **Data Model of Poverty in Papua Province with Semiparametric Regression through Weighted Penalized Spline** (supervised Anna Islamiyati).

**Background.** *In some cases, there is a pattern of relationship between the response variable and one or more known predictor variables and some others are unknown. This situation can be analyzed using a semiparametric regression approach. One of the estimators that can be used is a penalized spline with Weighted Least Square (WLS). This study applies this method to poverty data in Papua Province, with parametric predictor variables including average years of schooling and per capita expenditure, and nonparametric predictor variable being life expectancy. Aim.* This study aims to obtain parameter estimates and develop a model for poverty data in Papua Province with semiparametric regression through a weighted penalized spline. **Methods.** This study consists of two stages, the estimation of semiparametric regression parameters through a weighted penalized spline and the development of a poverty data model in Papua Province with semiparametric regression through a weighted penalized spline. **Results.** The estimation of semiparametric regression parameters through weighted penalized splines for the percentage of poverty in Papua Province in 2016-2021 are  $\hat{y} = X\hat{\beta} + Z\hat{\theta}$ . The model with the minimum Generalized Cross-Validation (GCV) value is the linear model with a penalization parameter of 0.3 and one knot point at 70.52. The model indicates that when life expectancy is below 70.52 years, the percentage of the population in poverty tends to increase by 0.480%. However, once life expectancy exceeds 70.52 years, the poverty rate significantly decreases. **Conclusion.** Thus, it can be concluded that the better semiparametric regression model through weighted penalized splines lies in the model that uses a one-knot point with a GCV value of 4.517.

**Keywords:** *Semiparametric Regression, Penalized Spline, Weighted Least Square, Poverty*

**DAFTAR ISTILAH**

---

<b>Istilah</b>	<b>Arti dan Penjelasan</b>
Estimator	Statistik sampel yang digunakan untuk menduga suatu parameter
Heteroskedastisitas	Kondisi dimana terdapat perbedaan variansi
Weighted Least Square	Metode pendugaan yang tidak memerlukan asumsi normalitas data serta memiliki sifat penduga yang konsisten
Orde	Derajat
Parameter penghalus	Pengontrol keseimbangan antara kemulusan kurva regresi dan kesesuaian fungsi terhadap data
Parameter	Nilai yang memperkirakan nilai populasi melalui suatu fungsi atau persamaan model
Variabilitas	Ukuran yang menyatakan seberapa besar nilai-nilai data berbeda atau menyimpang dari nilai ukuran pemusatannya
Variansi	Ukuran statistik yang menunjukkan seberapa jauh penyebaran data dari nilai rata-ratanya
Hipotesis	Suatu dugaan sementara terhadap masalah yang masih harus dibuktikan kebenarannya

---

## DAFTAR LAMBANG DAN SINGKATAN

Lambang/singkatan	Arti dan Penjelasan
$\beta$	Beta/parameter untuk komponen parametrik
$\theta$	Teta/parameter untuk komponen nonparametrik
$\varepsilon$	Nilai <i>error</i> acak pada pengamatan ke- $i$
$x$	Variabel prediktor untuk komponen parametrik
$z$	Variabel prediktor untuk komponen nonparametrik
$y$	Variabel respon
$k$	Titik knot
$W$	Pembobot
$\lambda$	Parameter penghalus
$q$	Orde
$H(\lambda)$	Matriks hat yang memuat parameter penghalus
$H_0$	Hipotesis nol
$H_1$	Hipotesis alternatif
$tr$	Trace/jumlah elemen pada diagonal utama matriks persegi
$i$	Pengamatan $i = 1, 2, \dots$
$n$	Jumlah pengamatan
$R^2$	Koefisien determinasi
$\sigma$	Variansi
$\alpha$	Tingkat signifikansi
$I$	Matriks identitas
PLS	<i>Penalized Least Square</i>
MSE	<i>Mean Square Error</i>
GCV	<i>Generalized Cross Validation</i>
WLS	<i>Weighted Least Square</i>

## DAFTAR ISI

	Halaman
<b>HALAMAN SAMPUL .....</b>	<b>i</b>
<b>HALAMAN JUDUL.....</b>	<b>i</b>
<b>HALAMAN PENGAJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....</b>	<b>vi</b>
<b>UCAPAN TERIMA KASIH.....</b>	<b>v</b>
<b>ABSTRAK.....</b>	<b>vi</b>
<b>ABSTRACT.....</b>	<b>vii</b>
<b>DAFTAR ISTILAH.....</b>	<b>viii</b>
<b>DAFTAR LAMBANG DAN SINGKATAN .....</b>	<b>ix</b>
<b>DAFTAR ISI .....</b>	<b>x</b>
<b>DAFTAR TABEL.....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Batasan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian .....	3
1.5 Kajian Teori.....	3
1.5.1 Regresi Parametrik .....	3
1.5.2 Regresi Nonparametrik .....	3
1.5.3 Regresi Semiparametrik.....	3
1.5.4 <i>Spline</i> .....	4
1.5.5 <i>Penalized Spline</i> .....	4
1.5.6 Regresi Semiparametrik dengan Pendekatan <i>Penalized Spline</i> ..	5
1.5.7 <i>Weighted Least Square</i> .....	6
1.5.8 <i>Generalized Cross Validation</i> .....	6
1.5.9 Kemiskinan .....	7

<b>BAB II METODE PENELITIAN .....</b>	<b>9</b>
2.1 Sumber Data.....	9
2.2 Variabel Penelitian .....	9
2.3 Metode Analisis.....	9
<b>BAB III HASIL DAN PEMBAHASAN .....</b>	<b>11</b>
3.1 Analisis Deskriptif.....	11
3.2 Estimasi Parameter Regresi Semiparametrik melalui Penalized Spline Terboboti.....	14
3.3 Model Persentase Kemiskinan di Provinsi Papua dengan Regresi Semiparametrik Melalui Penalized Spline Terboboti .....	16
<b>BAB IV KESIMPULAN DAN SARAN.....</b>	<b>21</b>
4.1 Kesimpulan .....	21
4.2 Saran .....	21
<b>DAFTAR PUSTAKA.....</b>	<b>23</b>
<b>LAMPIRAN .....</b>	<b>25</b>

**DAFTAR TABEL**

<b>Nomor Urut</b>		<b>Halaman</b>
1.	Deskriptif data kemiskinan di Provinsi Papua.....	11
2.	Nilai GCV minimum dengan satu titik knot pada variabel umur harapan hidup .....	16
3.	Nilai GCV minimum dengan dua titik knot pada variabel umur harapan hidup .....	16
4.	Nilai GCV minimum dengan tiga titik knot pada variabel umur harapan hidup .....	17
5.	Pemilihan orde, jumlah knot, dan parameter penghalus dengan membandingkan nilai GCV minimum .....	19

**DAFTAR GAMBAR**

<b>Nomor Urut</b>		<b>Halaman</b>
1.	<i>Scatter plot</i> variabel persentase penduduk miskin dan variabel umur harapan hidup.....	12
2.	<i>Scatter plot</i> variabel persentase penduduk miskin dan variabel rata-rata lama sekolah .....	12
3.	<i>Scatter plot</i> variabel persentase penduduk miskin dan variabel pengeluaran per kapita .....	13
4.	Kurva estimasi variabel umur harapan hidup .....	18
5.	Kurva estimasi variabel rata-rata lama sekolah.....	19
6.	Kurva estimasi variabel pengeluaran per kapita.....	19

**DAFTAR LAMPIRAN**

<b>Nomor Urut</b>		<b>Halaman</b>
1.	Data.....	29
2.	Uji Heteroskedastisitas .....	33
3.	<i>Output</i> Koefisien Determinasi .....	35

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Analisis regresi adalah analisis dalam statistika yang digunakan untuk melihat hubungan antara satu atau lebih variabel prediktor dengan variabel respon dalam sebuah penelitian. Terdapat dua pendekatan dalam analisis regresi, yaitu regresi parametrik dan regresi nonparametrik (Sanusi dkk., 2019). Pendekatan regresi parametrik digunakan jika bentuk hubungan diketahui seperti linier, kuadratik, kubik, eksponensial dan lainnya (Adrianingsih & Dani, 2021). Namun, jika bentuk hubungan antara variabel prediktor dan variabel respon tidak diketahui bentuknya atau tidak terdapat data historis yang lengkap mengenai bentuk hubungan tersebut, maka pendekatan regresi nonparametrik yang digunakan (Femadiyanti dkk., 2020).

*Spline* adalah salah satu metode dalam regresi nonparametrik yang terdiri dari potongan-potongan polinomial tersegmen yang digabungkan oleh beberapa titik knot untuk menghasilkan fungsi regresi yang sesuai dengan data dan mempertimbangkan kemulusan kurva (Zulaika dkk., 2024). Keunggulan lain yang dimiliki oleh *spline* ialah cenderung mencari sendiri estimasi data sesuai pergerakan pola data tersebut (Widyastuti dkk., 2021). Dalam penelitian yang dilakukan oleh Wulandary & Purnama (2020) mengenai pemodelan rata-rata lama sekolah dan pengeluaran per kapita di Indonesia menggunakan regresi kernel dan regresi *spline*, menunjukkan bahwa regresi *spline* lebih unggul dibandingkan regresi kernel.

Beberapa estimator yang umum digunakan dalam *spline* meliputi *truncated spline*, *B-spline*, dan *penalized spline*. Dari beberapa estimator *spline*, *penalized spline* merupakan estimator yang bersifat sederhana karena memungkinkan fleksibilitas dalam jumlah knot sehingga menghasilkan kurva yang lebih mulus. Estimator *penalized spline* diperoleh dengan meminimumkan fungsi *penalized least square* (PLS). PLS adalah suatu fungsi pendugaan yang digunakan untuk menggabungkan fungsi *least square* dengan kemulusan kurva (Kurniasari dkk., 2019). Dalam *penalized spline* digunakan salah satu metode *Weighted Least Square* (WLS) untuk memperbaiki estimasi parameter dalam model *spline* dengan mempertimbangkan variabilitas data (heteroskedastisitas). Pemilihan model estimator *penalized spline* dilakukan dengan mempertimbangkan orde, titik knot, dan parameter penghalus yang optimal menggunakan *Generalized Cross Validation* (GCV) minimum. Pada *penalized spline*, titik knot terletak pada nilai tunggal variabel prediktor, sehingga lebih efektif untuk data yang berjumlah besar (Insiro dkk., 2023). Alfairus dkk., (2021) melakukan penelitian menggunakan pendekatan nonparametrik dalam mengestimasi model kemiskinan di Sulawesi Selatan. Namun, pada kenyataannya seringkali ditemukan kasus dengan adanya pola hubungan antara variabel respon dengan satu atau beberapa variabel prediktor diketahui dan sebagian lain tidak diketahui. Oleh karena itu, diperlukan pengembangan penelitian yaitu pendekatan regresi semiparametrik.

Regresi semiparametrik adalah gabungan dari regresi parametrik dan regresi nonparametrik (Budiantara, 2014). Pendekatan regresi semiparametrik digunakan jika pola hubungan antara variabel respon dengan satu atau beberapa variabel prediktor diketahui dan ada pula bentuk pola hubungan dengan variabel prediktor yang lain tidak

diketahui (Khairunnisa dkk., 2020). Keunggulan dari metode ini terletak pada kemampuannya untuk mengidentifikasi pola hubungan yang belum diketahui dengan fleksibilitas dalam mengestimasi kurva regresi yang didukung oleh parameter yang digunakan dalam metode regresi nonparametrik. Beberapa model regresi nonparametrik meliputi kernel, polinomial lokal, *spline*, dan deret *fourier* (Abdy, 2019). Pada penelitian ini, model regresi semiparametrik melalui *penalized spline* dengan metode *weighted least square* diaplikasikan pada data persentase kemiskinan.

Kemiskinan merupakan salah satu tantangan utama yang dihadapi oleh banyak negara di dunia, termasuk Indonesia. Masalah ini tidak hanya mempengaruhi kesejahteraan sosial dan ekonomi masyarakat, tetapi juga memiliki dampak yang luas terhadap pertumbuhan ekonomi dan stabilitas sosial. Salah satu indikator yang digunakan untuk mengukur suatu kemiskinan adalah persentase penduduk miskin. Menurut Badan Pusat Statistik (2023), persentase kemiskinan di Provinsi Papua merupakan salah satu yang tertinggi di Indonesia. Beberapa faktor yang mempengaruhi persentase penduduk miskin seperti umur harapan hidup (Mariati dkk., 2022), rata-rata lama sekolah (Dani & Ni'matuzzahroh, 2021), dan pengeluaran per kapita (Siregar, 2023). Penelitian oleh Erlando dkk., (2022) menggunakan regresi semiparametrik dengan *spline truncated* untuk memodelkan tingkat pengangguran terbuka di Indonesia. Dzulhijjah (2021) menggunakan regresi semiparametrik dengan koefisien bervariasi pada data longitudinal menggunakan penaksir *b-spline*. Lisyanti (2024) menggunakan regresi semiparametrik dengan *smoothing spline*. Berdasarkan uraian tersebut, penulis akan melakukan pemodelan data kemiskinan di Provinsi Papua dengan regresi semiparametrik melalui *penalized spline* terboboti.

## 1.2 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data sekunder mengenai data kemiskinan di Papua tahun 2016 – 2021.
2. Pemilihan titik knot optimal menggunakan metode *Generalized Cross Validation* (GCV).
3. Jumlah titik knot yang dicobakan adalah satu, dua, dan tiga titik knot dengan orde yang dibatasi pada orde *linear*.
4. Pembobot yang digunakan adalah *Weighted Least Square*.

## 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Memperoleh hasil estimasi parameter regresi semiparametrik melalui *penalized spline* terboboti dalam memodelkan data kemiskinan di Provinsi Papua.
2. Memperoleh model data kemiskinan di Provinsi Papua dengan regresi semiparametrik melalui *penalized spline* terboboti.

## 1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Menambah wawasan dan pengetahuan mengenai regresi semiparametrik melalui *penalized spline* terboboti.
2. Memberikan informasi tentang regresi semiparametrik dalam memodelkan data kemiskinan di Provinsi Papua.
3. Memberikan informasi bagi pemerintah terkait data kemiskinan serta faktor yang mempengaruhinya dalam rangka pengambilan kebijakan masalah.

## 1.5 Kajian Teori

### 1.5.1 Regresi Parametrik

Regresi parametrik adalah metode statistika yang bertujuan untuk mengetahui adanya hubungan antara variabel respon dan variabel prediktor dengan asumsi bentuk kurva regresi yang telah diketahui sebelumnya. Regresi parametrik memiliki beberapa kelebihan ialah lebih sederhana, efisien, konsisten, dan estimator tidak bias (Setiawan dkk., 2022). Secara umum, model regresi parametrik ditulis sebagai berikut:

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

Keterangan:

- $y_i$  : variabel respon pada pengamatan ke  $-i$   
 $x_i$  : variabel prediktor pada pengamatan ke  $-i$   
 $f(x_i)$  : fungsi regresi yang telah diketahui bentuknya  
 $\varepsilon_i$  : *error* acak pada pengamatan ke  $-i$

### 1.5.2 Regresi Nonparametrik

Regresi nonparametrik adalah metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor dengan bentuk kurva regresi yang tidak diketahui, dengan kurva regresi hanya diasumsikan halus (*smooth*). Sehingga, regresi nonparametrik dianggap memiliki fleksibilitas yang tinggi karena data diharapkan untuk mencari sendiri bentuk estimasi kurva regresinya (Eubank, 1999).

Model umum regresi nonparametrik ditulis sebagai berikut:

$$y_i = g(z_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

Keterangan:

- $y_i$  : variabel respon pada pengamatan ke  $-i$   
 $g(z_i)$  : fungsi regresi pada pengamatan ke  $-i$  yang tidak diketahui  
 $z_i$  : variabel prediktor pada pengamatan ke  $-i$   
 $\varepsilon_i$  : *error* acak pada pengamatan ke  $-i$

### 1.5.3 Regresi Semiparametrik

Regresi semiparametrik adalah metode statistika yang digunakan ketika pola hubungan antara variabel respon dan variabel prediktor sebagian diketahui dan sebagian lagi tidak diketahui. Misalkan terdapat data berpasangan  $(y_i, x_i, z_i)$  dan model regresi

semiparametrik digunakan untuk mengasumsikan hubungan antara  $y_i, x_i$ , dan  $z_i$ . Model umum regresi semiparametrik ditulis sebagai berikut (Ruppert dkk., 2003):

$$y_i = f(x_i) + g(z_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3)$$

Keterangan:

$y_i$  : variabel respon pada pengamatan ke  $-i$

$x_i$  : variabel prediktor pada pengamatan ke  $-i$  untuk komponen parametrik

$z_i$  : variabel prediktor pada pengamatan ke  $-i$  untuk komponen nonparametrik

$f(x_i)$  : fungsi regresi pada pengamatan ke  $-i$  yang diketahui bentuknya

$g(z_i)$  : fungsi regresi pada pengamatan ke  $-i$  yang tidak diketahui bentuknya

$\varepsilon_i$  : error acak pada pengamatan ke  $-i$

### 1.5.4 Spline

*Spline* merupakan potongan-potongan polinomial tersegmen yang digabungkan oleh beberapa titik knot untuk menghasilkan fungsi regresi yang sesuai data dengan mempertimbangkan kemulusan kurva. Dalam fungsi *spline*, titik knot membentuk kurva yang dibentuk tersegmen. Titik di mana potongan-potongan digabungkan untuk menunjukkan perubahan pola perilaku fungsi *spline* pada berbagai interval disebut titik knot. Adanya titik knot membuat fungsi *spline* memiliki sifat fleksibilitas yang lebih baik daripada polinomial. Secara umum, fungsi *spline* dengan orde  $q$  dan titik knot  $k_1, k_2, \dots, k_r$  dapat dinyatakan sebagai berikut (Insiro dkk., 2023):

$$g(z_i) = \sum_{l=0}^q \theta_l z_i^l + \sum_{k=1}^r \theta_{q+k} (z_i - K_k)_+^q \quad (4)$$

Keterangan:

$\theta$  : parameter *spline*

$(z_i - k_k)_+^q$  : fungsi *truncated* yang menunjukkan segmentasi

### 1.5.5 Penalized Spline

Estimator *penalized spline* adalah salah satu estimator *spline* yang diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS). PLS adalah kriteria estimasi model antara *goodness of fit* dengan fungsi penalti (Islamiyati & Herdiani, 2019). Model dari *penalized least square* dapat ditulis sebagai berikut:

$$PLS = n^{-1} \sum_{i=1}^n (y_i - g(z_i))^2 + \lambda \sum_{k=1}^r \theta_{q+k}^2, \lambda \geq 0 \quad (5)$$

dengan  $n^{-1} \sum_{i=1}^n (y_i - g(z_i))^2$  adalah ukuran standar dari kesesuaian terhadap data yang terdiri dari *least square*,  $\sum_{h=1}^r \theta_{q+h}^2$  adalah ukuran kemulusan,  $\lambda$  adalah suatu parameter penghalus prediktor,  $k$  adalah jumlah knot, dan  $q$  adalah orde polinomial.

Untuk meminimumkan fungsi PLS pada Persamaan (5) dapat diubah ke dalam bentuk matriks sebagai berikut:

$$n^{-1} \sum_{i=1}^n (y_i - g(z_i))^2 = n^{-1} (\mathbf{y}^T \mathbf{y} - 2\boldsymbol{\theta}_j^T \mathbf{Z}_{ji}^T \mathbf{y} + \boldsymbol{\theta}_j^T \mathbf{Z}_{ji}^T \mathbf{Z}_{ji} \boldsymbol{\theta}_j) \quad (6)$$

$$\sum_k \theta_{q+k}^2 = \theta_{q+1}^2 + \theta_{q+2}^2 + \theta_{q+3}^2 + \dots + \theta_{q+k}^2 = \boldsymbol{\theta}_j^T \mathbf{D}_j \boldsymbol{\theta}_j \quad (7)$$

Jika diasumsikan matriks  $D_j$  adalah suatu matriks diagonal sebagai berikut:

$$D_j = \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I} \end{bmatrix}$$

dengan  $I$  adalah matriks identitas untuk  $\theta_{(q+1)}, \theta_{(q+2)}, \theta_{(q+3)}, \dots, \theta_{(q+k)}$ . Sehingga fungsi PLS dapat ditulis dalam bentuk matriks sebagai berikut:

$$PLS = n^{-1}(\mathbf{y}^T \mathbf{y} - 2\boldsymbol{\theta}_j^T \mathbf{Z}_{ji}^T \mathbf{y} + \boldsymbol{\theta}_j^T \mathbf{Z}_{ji}^T \mathbf{Z}_{ji} \boldsymbol{\theta}_j) + \lambda \boldsymbol{\theta}_j^T \mathbf{D}_j \boldsymbol{\theta}_j \quad (8)$$

Nilai  $\hat{\boldsymbol{\theta}}$  dapat diperoleh dengan meminumkan Persamaan (8) yaitu turunan pertama sama dengan nol sehingga diperoleh sebagai berikut:

$$\frac{\partial PLS}{\partial \boldsymbol{\theta}_j} = 0$$

$$n^{-1}(-2\mathbf{Z}_{ji}^T \mathbf{y} + 2\mathbf{Z}_{ji}^T \mathbf{Z}_{ji} \boldsymbol{\theta}_j + 2\lambda \boldsymbol{\theta}_j^T \mathbf{D}_j \boldsymbol{\theta}_j) = 0$$

$$\hat{\boldsymbol{\theta}}_j = (\mathbf{Z}_{ji}^T \mathbf{Z}_{ji} + n\lambda_j \mathbf{D}_j)^{-1} \mathbf{Z}_{ji}^T \mathbf{y} \quad (9)$$

Kemudian, substitusikan Persamaan (9) pada persamaan  $\hat{\mathbf{y}} = \mathbf{C}\hat{\boldsymbol{\theta}}$  sehingga diperoleh persamaan sebagai berikut:

$$\hat{\mathbf{y}} = \mathbf{Z}_j (\mathbf{Z}_{ji}^T \mathbf{Z}_{ji} + n\lambda_j \mathbf{D}_j)^{-1} \mathbf{Z}_{ji}^T \mathbf{y}$$

$$\hat{\mathbf{y}} = \mathbf{H}(\lambda_j) \mathbf{y} \quad (10)$$

### 1.5.6 Regresi Semiparametrik dengan Pendekatan *Penalized Spline*

Diberikan data berpasangan  $(y_i, x_i, z_i)$  untuk  $i = 1, 2, \dots, n$  dengan  $n$  menyatakan banyaknya pengamatan untuk setiap variabel prediktor,  $p$  menyatakan banyaknya variabel prediktor untuk komponen parametrik,  $v$  menyatakan banyaknya variabel prediktor untuk komponen nonparametrik. Model regresi semiparametrik *penalized spline* dapat ditulis sebagai berikut:

$$y_i = \sum_{s=0}^p \beta_s x_{si} + \sum_{t=1}^v g_t(z_{ti}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (11)$$

Keterangan:

$x_{si}$  : variabel prediktor pada pengamatan ke- $i$  untuk komponen parametrik

$\beta_s$  : parameter untuk komponen parametrik

$z_{ti}$  : variabel prediktor pada pengamatan ke- $i$  untuk komponen nonparametrik

$s$  :  $1, 2, \dots, p$

$t$  :  $1, 2, \dots, v$

Berdasarkan Persamaan (11), fungsi semiparametrik dengan estimator *penalized spline* berorde  $q$  dan titik knot  $k$  dapat ditulis sebagai berikut:

$$g_t(z_{ti}) = \sum_{l=0}^q \theta_k z_i^l + \sum_{k=1}^r \theta_{q+k} (z_i - k_k)_+^q \quad (12)$$

dengan fungsi *truncated* sebagai berikut:

$$(z_i - k_k)_+^q = \begin{cases} (z_i - k_k)^q & ; z_i \geq k_k \\ 0 & ; z_i < k_k \end{cases}$$

Dalam bentuk matriks, Persamaan (11) dapat dituliskan sebagai berikut:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (13)$$

Keterangan:

- $y$  : vektor variabel respon  
 $X$  : matriks variabel prediktor untuk komponen parametrik  
 $\beta$  : vektor parameter untuk komponen parametrik  
 $Z$  : matriks variabel prediktor untuk komponen nonparametrik  
 $\theta$  : vektor parameter untuk komponen nonparametrik  
 $\varepsilon$  : vektor *error*

### 1.5.7 Weighted Least Square

*Weighted Least Square* (WLS) adalah metode pendugaan yang tidak memerlukan asumsi normalitas data serta memiliki sifat penduga yang konsisten. Metode WLS hampir sama dengan *Ordinary Least Square* (OLS) yakni meminimumkan jumlah kuadrat *error* yang merupakan selisih antara nilai pengamatan variabel tak bias dengan nilai estimasinya. Namun, ketika estimasi OLS dilakukan terdapat salah satu atau lebih asumsi yang tidak terpenuhi, sehingga estimasi yang diperoleh tidak memenuhi sifat *Best Linear Unbiased Estimator* (BLUE). Oleh karena itu, metode *Weighted Least Square* (WLS) dapat dijadikan alternatif dalam melakukan estimasi parameter dengan memberikan pembobot pada model regresi (Dewi, 2021). Berdasarkan Persamaan (13), metode WLS dapat dirumuskan sebagai berikut:

$$\min\{\varepsilon^T W \varepsilon\} = \{(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\theta)^T \mathbf{W}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\theta)\} \quad (14)$$

dengan  $W$  adalah matriks pembobot dan  $\sigma_i^2$  sebagai pembobot yang dapat dinyatakan sebagai berikut:

$$W = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

### 1.5.8 Generalized Cross Validation

Pemilihan model *penalized spline* dilakukan dengan mempertimbangkan orde, jumlah knot, dan parameter penghalus yang nilai optimalnya menggunakan *Generalized Cross Validation* (GCV) minimum. Orde merujuk pada derajat polinomial yang digunakan pada setiap segmen, menunjukkan sejauh mana model dapat menyesuaikan dengan pola data. Parameter penghalus ( $\lambda$ ) merupakan pengontrol keseimbangan antara kemulusan kurva regresi dan kesesuaian fungsi terhadap data. Jika  $\lambda$  besar maka fungsi taksirannya akan semakin halus, sedangkan jika  $\lambda$  kecil maka fungsi taksiran yang didapat akan semakin besar atau fungsi tersebut menjadi semakin fluktuatif. Oleh karena itu, pemilihan parameter penghalus diharapkan memilih nilai  $\lambda$  yang optimal. Hasil fungsi taksiran atau estimasi fungsi regresi yang baik yaitu tidak terlalu halus serta tidak terlalu kasar (Sriliana dkk., 2020).

Jumlah knot adalah banyaknya titik knot atau banyaknya titik perubahan perilaku fungsi pada interval yang berbeda. Pemilihan jumlah knot yang sesuai berdasarkan nilai  $\lambda$  dan orde optimal. Titik knot adalah titik di mana potongan-potongan digabungkan untuk menunjukkan perubahan pola perilaku fungsi *spline* pada berbagai interval. Pada

*penalized spline*, titik knot terletak pada nilai *unique* (tunggal) variabel prediktor  $\{z_i\}_{i=1}^n$ , berarti hanya satu nilai yang dihitung jika terdapat beberapa nilai variabel prediktor yang sama.

Nilai GCV minimum akan dihitung untuk banyak titik knot  $k = 1, 2, \dots$  hingga banyak knot tertentu yang dicobakan. Maksimal jumlah knot yang dihitung dengan ketentuan  $k < (n_{unique} - q - 1)$ , dimana  $n_{unique}$  adalah banyaknya nilai *unique* (tunggal) dari variabel prediktor dan  $q$  adalah orde polinomial (Ruppert dkk., 2003).

Metode GCV dapat didefinisikan sebagai berikut:

$$GCV(\lambda) = \frac{MSE}{\left[1 - \frac{1}{n} \text{tr}[H(\lambda)]\right]^2} = \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left[1 - \frac{1}{n} \text{tr}[H(\lambda)]\right]^2} \quad (15)$$

Keterangan:

- $y_i$  : Variabel respon pada pengamatan ke- $i$
- $\hat{y}_i$  : Nilai estimasi variabel respon pada pengamatan ke- $i$
- $MSE$  : *Mean Square Error*
- $\lambda$  : Parameter penghalus
- $H(\lambda)$  : Matriks hat yang memuat parameter penghalus

### 1.5.9 Kemiskinan

Kemiskinan adalah ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Penduduk dikategorikan sebagai penduduk miskin apabila memiliki pengeluaran per kapita per bulan dibawah garis kemiskinan (Badan Pusat Statistik, 2023). Persentase penduduk miskin adalah persentase penduduk yang berada dibawah garis kemiskinan. Adapun beberapa faktor yang mempengaruhi adalah sebagai berikut:

#### 1. Umur harapan hidup

Umur harapan hidup (UHH) didefinisikan sebagai rata-rata perkiraan lama waktu yang dapat dijalani oleh seseorang selama hidupnya. Untuk menghitung umur harapan hidup adalah hasil bagi dari jumlah umur orang meninggal dibagi dengan jumlah orang yang meninggal.

#### 2. Rata-rata lama sekolah (RLS)

Rata-rata lama sekolah (RLS) adalah rata-rata lama jumlah tahun yang digunakan oleh penduduk usia 15 tahun ke atas dalam menjalani pendidikan formal. Adapun rumus dari RLS adalah sebagai berikut:

$$RLS = \frac{1}{n} \times \sum_{i=1}^n x_1$$

#### 3. Pengeluaran per kapita

Pengeluaran per kapita adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan baik yang berasal dari pembelian maupun produksi sendiri dibagi dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut. Pengeluaran dapat mengukur kemiskinan dari ketidakmampuan ekonomi untuk

memenuhi kebutuhan dasar pangan dan non pangan. Oleh karena itu, penduduk yang rata-rata pengeluaran per kapita per bulannya berada di bawah garis kemiskinan merupakan penduduk miskin. Kemiskinan menurun ketika pengeluaran per kapita tinggi, dan sebaliknya, kemiskinan meningkat ketika pengeluaran per kapita rendah.

## BAB II METODE PENELITIAN

### 2.1 Sumber Data

Data yang digunakan adalah data sekunder yang diperoleh dari *website* Badan Pusat Statistik (BPS) Papua pada tanggal 5 Juli 2024. Data yang digunakan adalah data kemiskinan di Papua tahun 2016 – 2021 dengan jumlah data sebanyak 174 data dan terlampir pada Lampiran 1.

### 2.2 Variabel Penelitian

Pada penelitian ini terdapat dua variabel penelitian yaitu variabel respon ( $Y$ ) dan variabel prediktor ( $X$ ). Adapun variabel penelitian yang digunakan adalah sebagai berikut

a. Variabel Respon

Variabel respon dalam penelitian ini adalah persentase penduduk miskin yang menggambarkan persentase penduduk yang berada dibawah garis kemiskinan.

b. Variabel Prediktor

Variabel prediktor dalam penelitian ini terdiri atas 2 variabel prediktor berpola parametrik dan 1 variabel prediktor berpola nonparametrik. Adapun variabel yang digunakan adalah sebagai berikut

1. Umur harapan hidup (UHH) yang disimbolkan dengan ( $z_1$ ), yaitu rata-rata perkiraan lama waktu yang dapat dijalani oleh seseorang selama hidupnya.
2. Rata-rata lama sekolah (RLS) yang disimbolkan dengan ( $x_1$ ), yaitu rata-rata lama jumlah tahun yang dihabiskan oleh penduduk usia 15 tahun ke atas dalam menjalani pendidikan formal.
3. Pengeluaran per kapita yang disimbolkan dengan ( $x_2$ ), yaitu biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga.

### 2.3 Metode Analisis

Langkah-langkah yang dilakukan untuk menganalisis data pada penelitian adalah sebagai berikut:

1. Melakukan statistika deskriptif untuk setiap variabel.
2. Membuat *scatter plot* antara variabel respon dengan variabel prediktor.
3. Menentukan variabel komponen parametrik dan komponen nonparametrik berdasarkan pola data dan koefisien determinasi antara variabel respon dengan variabel prediktor.
4. Memperoleh estimasi parameter regresi semiparametrik melalui *penalized spline* terboboti dalam memodelkan data kemiskinan di Provinsi Papua.
  - a. Diberikan data kemiskinan di Provinsi Papua dengan jumlah data 174 data, 1 variabel prediktor nonparametrik dan 2 variabel prediktor parametrik ( $y_i, x_{1i}, x_{2i}, z_i$ ) yang mengikuti model regresi semiparametrik *penalized spline* pada Persamaan (11).

- b. Menguraikan persamaan regresi semiparametrik *penalized spline* berdasarkan data kemiskinan di Provinsi Papua. Kemudian menyatakan dalam matriks berdasarkan Persamaan (13).
  - c. Mengestimasi model dengan meminimumkan kriteria *Penalized Weighted Least Square* (PWLS).
  - d. Mengestimasi nilai dugaan  $\hat{\beta}$  dengan mendiferensiasikan fungsi  $L$  terhadap  $\beta$ .
  - e. Mengestimasi nilai dugaan  $\hat{\theta}$  dengan mendiferensiasikan fungsi  $L$  terhadap  $\theta$ .
5. Menentukan jumlah knot dan parameter penghalus yang optimal berdasarkan nilai GCV minimum.
  6. Memodelkan data kemiskinan di Provinsi Papua dengan model regresi semiparametrik melalui *penalized spline* terboboti.
  7. Membuat kurva estimasi untuk setiap variabel prediktor.
  8. Melakukan interpretasi model dan menarik kesimpulan.