

**ESTIMASI INTERVAL KEPERCAYAAN
PARAMETER MODEL TINGKAT KEMISKINAN
MELALUI REGRESI SEMIPARAMETRIK *SPLINE*
UNTUK DATA LONGITUDINAL**

SKRIPSI



NUR AISYAH

H051191023

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

2024



**ESTIMASI INTERVAL KEPERCAYAAN
PARAMETER MODEL TINGKAT KEMISKINAN
MELALUI REGRESI SEMIPARAMETRIK *SPLINE*
UNTUK DATA LONGITUDINAL**

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan
Ilmu Pengetahuan Alam Universitas Hasanuddin**

**NUR AISYAH
H051191023**

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

**MAKASSAR
JANUARI 2024**



LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

**ESTIMASI INTERVAL KEPERCAYAAN PARAMETER MODEL
TINGKAT KEMISKINAN MELALUI REGRESI SEMIPARAMETRIK
SPLINE UNTUK DATA LONGITUDINAL**

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 22 Januari 2024



Nur Aisvah
NIM H051191023



**ESTIMASI INTERVAL KEPERCAYAAN
PARAMETER MODEL TINGKAT KEMISKINAN
MELALUI REGRESI SEMIPARAMETRIK *SPLINE*
UNTUK DATA LONGITUDINAL**

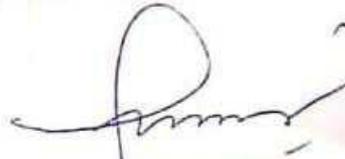
Disetujui Oleh:

Pembimbing Utama



Dr. Anna Islamiyati, S.Si., M.Si.
NIP. 197708082005012002

Pembimbing Pertama



Dr. Nirwan, M.Si.
NIP. 196303061987021002

Ketua Program Studi



Pada 22 Januari 2024



HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Nur Aisyah
NIM : H051191023
Program Studi : Statistika
Judul Skripsi : Estimasi Interval Kepercayaan Parameter Model Tingkat Kemiskinan Melalui Regresi Semiparametrik *Spline* Untuk Data Longitudinal

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Dr. Anna Islamiyati, S.Si., M.Si.

(.....)

2. Sekretaris : Dr. Nirwan, M.Si.

(.....)

3. Anggota : Siswanto, S.Si., M.Si.

(.....)

4. Anggota : Andi Kresna Jaya, S.Si., M.Si

(.....)

Ditetapkan di : Makassar

Tanggal : 22 Januari 2024



KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa Sallam* beserta keluarga dan para sahabatnya. *Alhamdulillahirobbil'alamin*, berkat nikmat kemudahan dan pertolongan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penulis dapat menyelesaikan skripsi yang berjudul “**Estimasi Interval Kepercayaan Parameter Model Tingkat Kemiskinan Melalui Regresi Semiparametrik Spline Untuk Data Longitudinal**” yang disusun sebagai salah satu syarat akademik untuk memperoleh gelar sarjana pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyelesaian skripsi ini tidak lepas dari bantuan dan dorongan dari berbagai pihak yang senantiasa turut membantu dalam bentuk moril maupun materil sehingga dengan segala keterbatasan kemampuan dan pengetahuan, penulis dapat menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis, Ayahanda **Paruddin** dan Ibunda **Juana** yang berjuang dalam memberikan dukungan serta petunjuk dalam perjalanan panjang perjuangan penulis, pengorbanan luar biasa, limpahan cinta dan kasih sayang, kesabaran hati, serta dengan ikhlas telah menemani setiap langkah penulis dengan doa dan restu mulianya. Ucapan terima kasih juga penulis haturkan kepada kakak tersayang penulis **Nur Aeynah** dan Adik-adik tersayang penulis **Ahmad Hidayat Fatli** dan **Muh. Yusuf Dirga** yang senantiasa memberikan semangat, dukungan moril dan doa terbaiknya untuk penulis dalam menyelesaikan tugas akhir ini, serta kepada keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan

penulis juga ucapkan kepada:

Prof. Dr. Ir. Jamaluddin Jompa, M.Sc., selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.



2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.**, selaku Ketua Departemen Statistika dan selaku Pembimbing Utama penulis dengan penuh kesabaran meluangkan waktu dan pengertiannya dalam membimbing, memberikan arahan, dorongan semangat dan motivasi, serta memahami kondisi masa sulit penulis dari awal hingga selesainya penulisan skripsi ini. Serta segenap dosen pengajar dan staf Departemen Statistika yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Bapak Dr. Nirwan, M.Si.**, selaku Pembimbing Pertama penulis dengan penuh kesabaran meluangkan waktu dan pengertiannya dalam membimbing, memberikan arahan, dorongan semangat dan motivasi, serta memahami kondisi masa sulit penulis dari awal hingga selesainya penulisan skripsi ini.
5. **Bapak Siswanto, S.Si., M.Si.**, selaku Penguji Pertama sekaligus Penasihat Akademik penulis yang senantiasa memberikan masukan dan arahan perihal akademik. Selain dari itu, beliau dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan masukan serta arahan sehingga menjadikan skripsi ini lebih baik.
6. **Bapak Andi Kresna Jaya, S.Si., M.Si.**, selaku Penguji Pertama penulis dengan penuh kesabaran telah meluangkan waktu dan pemikirannya untuk senantiasa memberikan masukan serta arahan yang membangun kepada penulis sehingga menjadikan skripsi ini lebih baik.
7. Teman-teman *Mahasiswi Santuy* **Amalia Mentari Djalumang, Andi Apridhani Mattalatta, Evi Sagita, Inda Muthmainna, Melinda Sari Wiryia Putri Laode, Muharti Apriana** dan **Rahmah Ningsih Dwika Julia**. Terima kasih telah menemani penulis sejak menjadi mahasiswi baru dan tetap menemani penulis hingga menyelesaikan masa studi.
8. Teman-teman **Andi Muhammad Rajab, Muhammad Syamsul Bahri, Al Hikmah** dan **Seli Lisnayati** dan **Vinaya Rifqi Anandari**. Terima kasih atas kebersamaan, kebahagiaan, bantuan, serta kebaikannya terhadap penulis.



9. Teman-teman KKN Gelombang 108 Pulau Samatellu Lompo .
10. Teman seperjuangan di Statistika 2019, **Muliana, Yasmin Pratiwi, Nurazizah, Nur Syahfika, Mayashari, Muhammad Faturrahman, Sapriadi Rasyid, Ahmaf Firdaus Irianto, Muhammad Rayhan Rifaldi, Muhammad Ferdiansyah, Eplysius Febriano Ndarung, Arief Rahman Nur, Alfiya Salsabila, Fadhila Febriyanti Najamuddin, Diah Lestari, Muhammad Nur Iskandar Zulkarnain, Muhammad Yusran, Ummul Auliyah Syam dan Wahyu Dwi Rahmawati.** Terima kasih atas kebaikan dan bantuannya kepada penulis sehingga dapat menuntaskan masa penyusunan skripsi dengan baik.
11. Kepada semua teman-teman di angkatan Statistika 2019 yang telah mengisi masa perkuliahan penulis dengan banyak sekali cerita yang akan terus terkenang dalam ingatan penulis.
12. Diri saya sendiri, yang telah mampu bertahan dan tidak berhenti dalam mengerjakan tugas akhir ini. Terimakasih karena selalu berpikir positif ketika keadaan sempat tidak berpihak, dan selalu berusaha mempercayai diri sendiri.
13. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan, partisipasi, dan apresiasi yang diberikan kepada penulis.

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

Makassar, 22 Januari 2024



Nur Asyiah



**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Nur Aisyah
NIM : H051191023
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

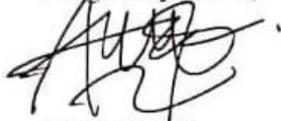
**“Estimasi Interval Kepercayaan Parameter Model Tingkat Kemiskinan
Melalui Regresi Semiparametrik Spline Untuk Data Longitudinal”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar tanggal 22 Januari 2024.

Yang menyatakan,


(Nur Aisyah)



ABSTRAK

Estimasi parameter dapat dilakukan dengan melalui estimasi titik dan estimasi interval kepercayaan. Interval kepercayaan pada regresi parametrik dan nonparametrik telah banyak dilakukan, namun untuk regresi semiparametrik belum banyak dilakukan apalagi pada data longitudinal. Pendekatan dalam regresi semiparametrik yang digunakan pada penelitian ini yaitu *spline truncated*. Kemudian untuk mendapatkan interval kepercayaan digunakan pendekatan *pivotal quantity*. Penelitian ini bertujuan untuk mendapatkan estimator interval kepercayaan pada regresi semiparametrik *spline* pada data longitudinal dan mengaplikasikannya pada data tingkat kemiskinan di Provinsi Papua. Dari hasil penelitian didapatkan interval kepercayaan regresi semiparametrik *spline* ketika σ^2 tidak diketahui. Kemudian diperoleh model terbaik regresi semiparametrik *spline truncated* pada persentase penduduk miskin Provinsi Papua dengan menggunakan tiga titik knot dan memiliki nilai R^2 sebesar 90.779% dan MSE sebesar 11.423. Berdasarkan hasil inferensia statistik dengan menggunakan nilai interval kepercayaan untuk parameter model regresi semiparametrik *spline truncated* untuk data longitudinal yang diterapkan pada data tingkat kemiskinan Provinsi Papua ditunjukkan bahwa dari 203 parameter terdapat 145 parameter yang signifikan dan 58 parameter yang tidak signifikan.

Kata Kunci: Regresi Semiparametrik, *Spline*, Interval Kepercayaan, Data Longitudinal, Tingkat Kemiskinan.



ABSTRACT

Parameter estimation can be done through point estimation and confidence interval estimation. Confidence intervals in parametric and non-parametric regression have been widely done, but for semiparametric regression has not been done much especially on longitudinal data. The approach in semiparametric regression used in this study is truncated spline. Then to get the confidence interval a pivotal quantity approach is used. This study aims to obtain the confidence interval estimator in semiparametric spline regression on longitudinal data and apply it to poverty rate data in Papua Province. From the research results, the confidence interval of semiparametric spline regression is obtained when σ^2 is unknown. Then obtained the best model of truncated spline semiparametric regression on the percentage of poor people in Papua Province using three knot points and has an R^2 value of 90.779% and MSE of 11.423. Based on the results of statistical inference using the confidence interval value for the parameters of the semiparametric spline truncated regression model for longitudinal data applied to the poverty rate data of Papua Province, it is shown that out of 203 parameters, 145 parameters are significant and 58 parameters are not significant.

Keywords: *Semi-parametric Regression, Spline, Confidence Interval, Longitudinal Data, Poverty Levels.*



DAFTAR ISI

HALAMAN JUDUL ii

LEMBAR PERNYATAAN KEOTENTIKANError! Bookmark not defined.

HALAMAN PERSETUJUAN PEMBIMBING iv

HALAMAN PENGESAHANError! Bookmark not defined.

KATA PENGANTAR..... v

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI..... vii

ABSTRAK viii

ABSTRACT ix

DAFTAR ISI..... x

DAFTAR TABEL..... xi

DAFTAR GAMBAR..... xii

DAFTAR LAMPIRAN xiii

BAB I PENDAHULUAN..... 1

1.1 Latar Belakang..... 1

1.2 Rumusan Masalah..... 3

1.3 Batasan Masalah 3

1.4 Tujuan Penelitian 4

1.5 Manfaat Penelitian..... 4

BAB II TINJAUAN PUSTAKA..... 5

2.1 Analisis Regresi 5

 2.1.1 Regresi Parametrik 5

 2.1.2 Regresi Nonparametrik 5

 2.1.3 Regresi Semiparametrik..... 6

2.2 *Spline* 6

2.3 Regresi Semiparametrik dengan *Spline Truncated*..... 7

2.4 Data Longitudinal 7

2.5 Regresi Semiparametrik dengan *Spline Truncated* pada Data Longitudinal. 8

2.6 *Weighted Least Square* 9

 Pilihan Titik Knot Optimal 10

 Interval Kepercayaan Parameter 10

 Persentase Penduduk Miskin 11



BAB III METODOLOGI PENELITIAN	13
3.1 Data.....	13
3.2 Metode Analisis	13
3.3 Diagram Alir.....	16
BAB IV HASIL DAN PEMBAHASAN.....	17
4.1 Konstruksi Interval Kepercayaan Regresi Semiparametrik <i>Spline Truncated</i> pada Data Longitudinal	17
4.2 Aplikasi Interval Kepercayaan Regresi Semiparametrik <i>Spline Truncated</i> pada Data Tingkat Kemiskinan di Provinsi Papua	29
BAB V KESIMPULAN DAN SARAN	39
5.1 Kesimpulan.....	39
5.2 Saran	39
DAFTAR PUSTAKA	40



DAFTAR TABEL

Tabel 2.1 Struktur Data Longitudinal untuk Regresi Semiparametrik.....	8
Tabel 3.1 Variabel Respon dan Variabel Prediktor.....	13
Tabel 4.1 Statistik Deskriptif.....	30
Tabel 4.2 Nilai GCV dengan Dua Titik Knot	34
Tabel 4.3 Nilai GCV dengan Tiga Titik Knot.....	34
Tabel 4.4 Perbandingan Nilai GCV Minimum.....	35
Tabel 4.5 Interval Kepercayaan untuk Parameter	37



DAFTAR GAMBAR

Gambar 4.1 *Scatter Plot* Persentase Penduduk Miskin dan Angka Harapan Hidup 31

Gambar 4.2 *Scatter Plot* Persentase Penduduk Miskin dan Rata-rata Lama Sekolah..... 32

Gambar 4.3 *Scatter Plot* Persentase Penduduk Miskin dan Pengeluaran Per kapita..... 32

Gambar 4.4 *Scatter Plot* Persentase Penduduk Miskin dan Angka Harapan Hidup 37



BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan metode yang digunakan untuk mengetahui hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Terdapat 3 jenis pendekatan dalam analisis regresi yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik (Sudiarsa dkk., 2015). Dalam regresi parametrik bentuk kurva regresi diasumsikan diketahui. Berbeda dengan pendekatan parametrik, dalam regresi nonparametrik bentuk kurva regresi diasumsikan tidak diketahui. Kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat di dalam suatu ruang fungsi tertentu (Eubank, 1988). Adapun regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik. Model regresi semiparametrik digunakan apabila variabel respon memiliki pola hubungan yang diketahui dengan salah satu atau beberapa variabel prediktor, akan tetapi dengan variabel prediktor yang lain tidak diketahui dengan pasti pola hubungannya (Budiantara, 2009).

Pada kasus riil, seringkali dijumpai fenomena masing-masing variabel prediktor memiliki pola hubungan yang berbeda-beda. Maka dari itu, jika pola hubungan variabel prediktor dan variabel respon diketahui maka akan dilakukan pendekatan regresi parametrik sedangkan untuk pola hubungan variabel prediktor dan variabel respon yang tidak diketahui maka akan dilakukan pendekatan menggunakan regresi nonparametrik (Adrianingsih & Andrea, 2021). Terdapat beberapa metode untuk mengestimasi kurva regresi dengan menggunakan pendekatan nonparametrik yaitu *kernel*, *spline*, polinomial lokal, wavelet dan deret fourier (Eubank, 1999). Diantara metode pendekatan tersebut, regresi nonparametrik dengan pendekatan *spline* merupakan metode yang sering digunakan. *Spline* memiliki fleksibilitas yang tinggi serta memiliki kemampuan yang sangat baik dalam menangani data yang perilakunya berubah-ubah pada sub-sub internal tertentu (Budiantara, 2009). Pada penelitian yang dilakukan oleh

kk (2022) dengan membandingkan regresi *spline* dan *kernel* diperoleh lebih baik dalam memodelkan angka kematian bayi di Jawa Barat. Salah satu metode yang digunakan dalam *spline* adalah *spline truncated*.



Spline truncated merupakan model regresi yang mempunyai interpretasi statistik visual yang sangat baik. Selain itu, kelebihan *spline truncated* adalah mampu mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus. Titik knot merupakan perpaduan bersama yang menunjukkan pola perilaku fungsi *spline truncated* pada selang yang berbeda (Hardle, 1994). Pratama dkk (2022) telah melakukan pemodelan regresi nonparametrik dengan menggunakan pendekatan *spline truncated* dan memperoleh faktor-faktor yang berpengaruh secara signifikan pada kasus *tuberculosis*. Estimasi kurva regresi nonparametrik *spline truncated* dapat dilakukan dengan memilih parameter *smoothing*, yaitu orde, banyaknya titik knot, dan titik knot. Pemilihan parameter *smoothing* yang optimal dilakukan dengan menggunakan kriteria *Generalized Cross Validation* (GCV) yaitu dengan memilih nilai GCV yang minimum. GCV digunakan karena memiliki sifat optimal asimtotik yang tidak dimiliki metode lain (Wahba, 1990).

Estimasi parameter dapat dilakukan dengan melalui estimasi titik dan estimasi interval kepercayaan. Jika hasil dugaan yang berupa titik lebih rendah dibandingkan dengan galat dugaannya akan memberikan hasil yang tidak akan baik, maka dari itu pendugaan parameter dapat dilakukan dengan pendugaan melalui interval kepercayaan (Marzuki dkk, 2010). Estimasi interval mendekati suatu nilai parameter dengan menggunakan dua titik yang memungkinkan untuk mengukur derajat kepercayaan terhadap ketelitian dari estimasi (Ma'unnah, 2016). Pada penelitian yang dilakukan oleh Hadijah (2022) pada data pasien demam berdarah *dengue* dengan mengestimasi interval kepercayaan parameter pada model regresi *penalized* kuantil *spline*. Penelitian dengan pendekatan nonparametrik banyak dilakukan pada bidang kesehatan namun juga dapat diaplikasikan pada bidang lain termasuk di bidang sosial dan ekonomi (Wu & Zhang, 2007). Pada penelitian ini, interval kepercayaan parameter model regresi semiparametrik *spline truncated* diaplikasikan pada data tingkat kemiskinan.

Kemiskinan yang menjadi tujuan utama *Millennium Development*

MDGs) kembali menjadi tujuan utama dalam *Sustainable Development* (SDGs) dan menjadikan penghapusan kemiskinan sebagai tujuan utama untuk menjamin keberlanjutan capaian MDGs (Bappenas, 2017). Salah



satu indikator yang digunakan untuk mengukur suatu kemiskinan adalah persentase penduduk miskin (BPS, 2023). Menurut data Badan Pusat Statistik (2023) provinsi dengan persentase kemiskinan tertinggi di Indonesia yaitu provinsi Papua. Pada provinsi Papua, terjadi peningkatan ataupun penurunan persentase penduduk miskin di tiap kabupaten/kota. Adanya perbedaan persentase penduduk miskin yang terjadi di setiap wilayah pada beberapa tahun sehingga perlu dilakukan penelitian yang memperhatikan lokasi dan efek waktu. Pada penelitian yang dilakukan oleh Umam dkk (2018) dengan menggunakan estimator *spline truncated* telah memperoleh interval kepercayaan parameter regresi nonparametrik pada data longitudinal dengan studi kasus data angka gini rasio di Indonesia. Astuti (2018) pada penelitiannya mengestimasi interval kepercayaan dengan menggunakan pendekatan *spline truncated* pada data longitudinal. Data longitudinal merupakan data yang dikumpulkan pada satu waktu tertentu dengan beberapa objek amatan yang terdiri dari beberapa periode waktu (Peter dkk, 2013). Pada pengamatan yang dilakukan secara berulang terhadap subjek mengakibatkan respon yang diperoleh satu dengan yang lainnya akan saling berhubungan (Diggle dkk, 2002).

Pada penelitian sebelumnya persentase penduduk miskin dipengaruhi oleh faktor seperti angka harapan hidup (Mariati dkk., 2022) rata-rata lama sekolah (Dani & Ni'matuzzahroh, 2020) dan pengeluaran per kapita (Muhtarom, 2016). Oleh karena itu, dalam penelitian ini akan dilakukan estimasi interval kepercayaan parameter untuk data longitudinal pada regresi semiparametrik dengan pendekatan *spline truncated*. Selain itu, untuk mengetahui faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Papua.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi interval kepercayaan pada regresi semiparametrik dengan menggunakan estimator *spline truncated* pada data longitudinal?
2. Bagaimana penerapan hasil estimasi interval kepercayaan parameter pada kemiskinan di Provinsi Papua dengan regresi semiparametrik menggunakan pendekatan *spline truncated*?



1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Konstruksi interval kepercayaan untuk parameter model regresi semiparametrik menggunakan model *spline truncated* linier.
2. Pendekatan yang digunakan untuk mengkonstruksi interval kepercayaan pada penelitian ini adalah pendekatan *pivotal quantity*.
3. Data yang digunakan adalah data yang diperoleh dari Badan Pusat Statistik Provinsi Papua (2014-2021).
4. Penentuan model regresi semiparametrik terbaik berdasarkan nilai GCV.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan estimator interval kepercayaan pada regresi semiparametrik dengan menggunakan estimator *spline truncated* pada data longitudinal.
2. Mendapatkan hasil estimasi interval kepercayaan parameter pada tingkat kemiskinan di Provinsi Papua dengan regresi semiparametrik menggunakan pendekatan *spline truncated*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Memberikan informasi tentang pemodelan regresi semiparametrik mengenai persentase penduduk miskin di provinsi Papua.
2. Memberikan informasi tentang analisis interval kepercayaan pada regresi semiparametrik berdasarkan estimator *spline truncated* pada data longitudinal.
3. Memberikan gagasan baru bagi pembaca dalam mengembangkan analisis regresi semiparametrik ketika mengestimasi interval kepercayaan dengan menggunakan estimator yang berbeda atau dengan kasus lain yang ditemukan dalam kehidupan sehari-hari.

Sebagai sumber informasi bagi Pemerintah provinsi Papua terkait persentase penduduk miskin serta faktor yang mempengaruhinya dalam langkah untuk memberikan kebijakan dalam permasalahan tersebut.



BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu metode statistika yang digunakan untuk mengukur ada atau tidaknya hubungan antara satu atau lebih variabel (Eubank, 1999). Secara umum, model analisis regresi yang paling sederhana memiliki bentuk Persamaan (2.1):

$$y = f(x_i) + \varepsilon \quad (2.1)$$

Dalam analisis regresi terdapat tiga model yaitu model regresi parametrik, model regresi nonparametrik, dan model regresi semiparametrik. Pada model regresi parametrik bentuk kurva regresi diasumsikan diketahui sedangkan regresi nonparametrik bentuk kurva regresi diasumsikan tidak diketahui. Kurva regresi nonparametrik hanya diasumsikan *smooth* (mulus) dalam arti termuat di dalam suatu fungsi tertentu. Bentuk estimasi dari data tanpa diketahui oleh faktor subyektifitas peneliti. Regresi semiparametrik digunakan jika di dalam suatu model regresi terdapat komponen parametrik dan nonparametrik (Budiantara, 2011).

2.1.1 Regresi Parametrik

Pendekatan regresi parametrik dipergunakan jika diasumsikan bahwa penyebaran data membentuk pola tertentu seperti: linier, kuadrat, kubik, maupun polinomial berderajat-p. Untuk dapat menggunakan model regresi parametrik terdapat asumsi tegas, yaitu bentuk kurva regresi diketahui (Eubank, 1988). Secara umum bentuk regresi parametrik linier ditulis pada Persamaan (2.2):

$$y_i = f(x_i) + \varepsilon_i \quad (2.2)$$

dengan y_i adalah nilai variabel respon dalam amatan ke-i, x_i adalah peubah bebas atau variabel prediktor dalam amatan ke-i, $f(x_i)$ adalah regresi yang telah diketahui bentuknya, ε_i adalah suku galat atau sisaan yang diasumsikan terdistribusi normal dan bersifat acak dengan nilai tengah 0 dan varians σ^2 (Yani dkk.,



2.1.2 Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metoda statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor. Apabila hubungan antara variabel respon dengan variabel prediktor tidak diketahui secara jelas polanya, atau tidak didapatkan informasi sebelumnya yang lengkap tentang bentuk pola data, maka digunakan pendekatan regresi nonparametrik. Pendekatan regresi nonparametrik digunakan untuk menduga kurva regresi yang tidak diketahui bentuk kurva regresinya dan tidak terikat pada asumsi tertentu seperti pada regresi parametrik. Persamaan regresi nonparametrik dinyatakan pada Persamaan (2.3):

$$y_i = g(z_i) + \varepsilon_i \tag{2.3}$$

y_i adalah variabel respon ke-i, $g(z_i)$ adalah kurva regresi yang tidak diketahui bentuk kurvanya dan ε_i adalah galat *random* yang diasumsikan independen dan identik dengan *mean* 0 dan varians σ^2 .

2.1.3 Regresi Semiparametrik

Regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik. Keadaan ketika variabel respon diketahui hubungannya dengan salah satu variabel prediktor, tetapi dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Menurut Dasty dkk (2019) model regresi semiparametrik seperti pada Persamaan (2.4):

$$y_i = f(x_i) + g(z_i) + \varepsilon_i \tag{2.4}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} g(z_{11}, z_{12}, \dots, z_{1r}) \\ g(z_{21}, z_{22}, \dots, z_{2r}) \\ \vdots \\ g(z_{n1}, z_{n2}, \dots, z_{nr}) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Estimasi fungsi regresi nonparametrik yang terdapat pada regresi semiparametrik dilakukan berdasarkan data pengamatan dengan menggunakan beberapa metode. Metode regresi nonparametrik yaitu *kernel*, *k-nearest neighbor*, *orthogonal*, dan *spline* (Hardle, 1994).

spline

spline merupakan salah satu teknik estimasi regresi nonparametrik yang



pertama kali dikembangkan oleh Whittaker pada tahun 1923 (Hardle, 1990). *spline* mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan (Eubank, 1988). Secara umum fungsi *spline* $f(z_i)$ berorde a dengan titik knot k_1, k_2, \dots, k_l dapat dinyatakan pada Persamaan (2.5):

$$f(z_i) = \sum_{k=1}^r \alpha_k z_{ik} + \sum_{u=1}^l \gamma_{a+u} (z_{ik} - k_{uk})_+^a \quad (2.5)$$

dengan $\sum_{k=1}^r \alpha_k z_{ik}$ merupakan komponen polinomial dan $\sum_{u=1}^l \gamma_{a+u} (z_{ik} - k_{uk})_+^a$ merupakan komponen *truncated* (potongan) (Polanagau dkk., 2019). Adapun Persamaan regresi nonparametrik *spline* dapat diperoleh dengan mensubstitusi Persamaan (2.3) ke Persamaan (2.5) sehingga diperoleh Persamaan (2.6):

$$y_i = \sum_{k=1}^r \alpha_k z_{ik} + \sum_{u=1}^l \gamma_{a+u} (z_{ik} - k_{uk})_+^a + \varepsilon_i \quad (2.6)$$

dengan α_k adalah nilai parameter model dan a merupakan orde *spline* (Purnaraga dkk., 2020).

2.3 Regresi Semiparametrik dengan Pendekatan *Spline Truncated*

Spline truncated merupakan salah satu pendekatan basis *spline* dalam regresi nonparametrik. Salah satu kelebihan dari *spline truncated* yaitu model ini mengikuti pola sesuai pergerakannya dengan titik knot (Astuti, 2017). *Spline truncated* memiliki kurva regresi dari hasil modifikasi fungsi *polynomial*. Diberikan data berpasangan (y_i, x_i, z_i) dengan $i = 1, 2, \dots, n$ dengan y_i adalah variabel respon sedangkan x_i adalah variabel prediktor yang mengikuti pola parametrik dan z_i adalah variabel prediktor yang mengikuti pola nonparametrik, model regresi semiparametrik *spline truncated* dengan q merupakan banyaknya variabel prediktor sehingga dapat ditulis menjadi Persamaan (2.7):

$$y_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^r \alpha_k z_i + \sum_{u=1}^k \gamma_{(u+a)k} (z_i - K_{uk})_+^a + \varepsilon_i \quad (2.7)$$

Keterangan:

y_i : Variabel respon pada data ke-i

β_0 : Parameter intersep untuk variabel komponen parametrik

β_1 : Parameter untuk variabel komponen parametrik

α_k : Parameter untuk variabel komponen parametrik pada data ke-i

$\gamma_{(u+a)k}$: Parameter untuk variabel komponen nonparametrik

ε_i : Variabel komponen nonparametrik pada data ke-i



γ : Parameter *truncated*

dengan $\alpha_k z_i$ merupakan komponen *truncated polynomial* yang mengikuti:

$$(z_i - K_{uk})_+^a = (z_{ijk} - K_{uk})_+^a = \begin{cases} (z_i - K_{uk})_+^a, & z_{ijk} \geq K_{uk} \\ 0, & z_{ijk} < K_{uk} \end{cases}$$

2.4 Data Longitudinal

Menurut Wu dan Zhang (2006) data longitudinal merupakan data yang diperoleh dari pengukuran berulang pada setiap subyek dalam kurun waktu berbeda. Dalam studi longitudinal, data dari individu dikumpulkan berulang dari waktu ke waktu sedangkan penelitian *cross-section* hanya mendapatkan satu data dari masing-masing individu, misalnya dengan satu waktu per subyek. Sebuah studi longitudinal, data dari subyek yang berbeda diasumsikan saling bebas, sedangkan data dari subyek yang sama terdapat kemungkinan saling berkorelasi. Perbedaan utama antara data longitudinal dan data *cross-section* adalah bahwa data longitudinal berkorelasi dalam satu subyek pengamatan dan saling bebas antar subyek pengamatan, sedangkan data *cross-section* saling bebas dalam subyek maupun antar subyek dalam pengamatan.

Tabel 2. 1 Struktur Data Longitudinal untuk Regresi Semiparametrik

Subyek (i)	Waktu (j)	Respon (y_{il}) $i = 1, 2, \dots, n;$ $j = 1, 2, \dots, m$	Prediktor (Parametrik) (x_1, x_2, \dots, x_q)				Prediktor (Nonparametrik) (z_1, z_2, \dots, z_r)			
Subyek 1										
	1	y_{11}	x_{111}	x_{112}	...	x_{11q}	z_{111}	z_{112}	...	z_{11r}
	2	y_{12}	x_{121}	x_{122}	...	x_{12q}	z_{121}	z_{122}	...	z_{12r}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	m	y_{1m}	x_{1m1}	x_{1m2}	...	x_{1mq}	z_{1m1}	z_{1m2}	...	z_{1mr}
Subyek 2										
	1	y_{21}	x_{211}	x_{212}	...	x_{21q}	z_{211}	z_{212}	...	z_{21r}
	2	y_{22}	x_{221}	x_{222}	...	x_{22q}	z_{221}	z_{222}	...	z_{22r}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	m	y_{2m}	x_{2m1}	x_{2m2}	...	x_{2mq}	z_{2m1}	z_{2m2}	...	z_{2mr}



Subyek n									
1	y_{n1}	x_{n11}	x_{n12}	...	x_{n1q}	z_{n11}	z_{n12}	...	z_{n1r}
2	y_{n2}	x_{n21}	x_{n22}	...	x_{n2q}	z_{n21}	z_{n22}	...	z_{n2r}
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
m	y_{nm}	x_{nm1}	x_{nm2}	...	x_{nmq}	z_{nm1}	z_{nm2}	...	z_{nmr}

2.5 Regresi Semiparametrik dengan Pendekatan *Spline Truncated* pada Data Longitudinal

Misalkan x_{ij} menyatakan pengamatan pada subyek ke- i waktu ke- j pada variabel prediktor yang mengikuti pola parametrik dan z_{ij} menyatakan pengamatan pada subyek ke- i dan waktu ke- j pada variabel prediktor yang mengikuti pola nonparametrik dengan $j = 1, 2, \dots, m$. Bentuk Persamaan regresi semiparametrik *spline truncated* untuk data longitudinal dapat ditulis pada Persamaan (2.8):

$$y_{ij} = \sum_{p=1}^q \beta_{ip} x_{ijp} + \sum_{k=1}^r \left(\alpha_{ik} z_{ijk} + \sum_{u=1}^l \gamma_{(u+\alpha)k} (z_{ijk} - K_{uk})_+^a \right) + \varepsilon_{ij} \quad (2.8)$$

dengan $\sum_{u=1}^l \gamma_{(u+\alpha)k} (z_{ijk} - K_{uk})_+^a$ merupakan komponen *truncated* yang memenuhi Persamaan (2.9):

$$(z_{ijk} - K_{uk})_+^a = \begin{cases} (z_{ijk} - K_{uk})_+^a, & z_{ijk} \geq K_{uk} \\ 0, & z_{ijk} < K_{uk} \end{cases} \quad (2.9)$$

dengan α_j dan γ_{ujk} dengan $u = 1, 2, \dots, k$ adalah parameter yang tidak diketahui (Wening dkk., 2020). Persamaan (2.8) dapat ditulis dalam bentuk matriks pada Persamaan (2.10):

$$y = X\beta + Z\eta + \varepsilon \quad (2.10)$$

dengan:

y : vektor variabel respon, berukuran $nm \times 1$

X : matriks variabel prediktor parametrik, berukuran $nm \times p$

β : vektor parameter regresi parametrik, berukuran $np \times 1$

Z : matriks variabel prediktor nonparametrik, berukuran $nm \times k$

η : vektor parameter regresi nonparametrik, berukuran $nk \times 1$

ε : vektor variabel galat, berukuran $nm \times 1$

ε berdistribusi normal identik dengan *mean* 0 dan varians $\sigma^2 \mathbf{I}$.



2.6 Weighted Least Square

Model regresi yang diterapkan pada data longitudinal memiliki dua asumsi mendasar. Asumsi pertama yaitu varians random galat dalam model diasumsikan homogen untuk setiap pengukuran berulang pada subjek dan asumsi kedua adalah matriks varians-kovarians dalam model diasumsikan diketahui (Budiantara, 2010). Metode *weighted least square* mengestimasi parameter dengan meminimalkan jumlah kuadrat antara observasi dan model (meminimumkan jumlah kesalahan kuadrat). Berdasarkan Persamaan (2.10), secara umum metode *weighted least square* untuk mengestimasi parameter dapat ditulis pada Persamaan (2.11) (Kuzairi dkk., 2022):

$$\min_{\beta, \eta} [R(\beta, \eta) = \min_{\beta, \eta} \{(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\eta)^T \mathbf{W}^{-1}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\eta)\} \quad (2.11)$$

dengan

$$\mathbf{W} = \begin{bmatrix} \frac{1}{n} & 0 & \dots & 0 \\ 0 & \frac{1}{n} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n} \end{bmatrix}$$

Berdasarkan metode *weighted least square* diperoleh estimator untuk parameter $\hat{\beta}$ dan $\hat{\eta}$ adalah

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \{ \mathbf{X}^T \mathbf{W}^{-1} \mathbf{y} - \mathbf{X}^T \mathbf{W}^{-1} \mathbf{Z} \hat{\eta} \} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{-1} (\mathbf{y} - \mathbf{Z} \hat{\eta}) \\ \hat{\eta} &= (\mathbf{Z}^T \mathbf{W}^{-1} \mathbf{Z})^{-1} \{ \mathbf{Z}^T \mathbf{W}^{-1} \mathbf{y} - \mathbf{Z}^T \mathbf{W}^{-1} \mathbf{X} \hat{\beta} \} = (\mathbf{Z}^T \mathbf{W}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) \end{aligned}$$

2.7 Pemilihan Titik Optimal

Titik knot merupakan titik yang terdapat pada perubahan pola perilaku fungsi. Salah satu metode yang biasa digunakan untuk memilih titik knot optimal adalah metode *Generalized Cross Validation* (GCV). Metode GCV mempunyai sifat optimal asimtotik. Model regresi *Spline* terbaik diperoleh dari titik knot optimal dengan melihat nilai GCV terkecil (Wahba, 1990). Pemilihan titik knot yang optimal dilakukan dengan melihat nilai GCV dari masing-masing orde dan

pada setiap subyek, lalu dipilih nilai GCV yang paling minimum. Fungsi at dituliskan dalam bentuk Persamaan (2.12) (Pratiwi, 2020):

$$GCV_i(K) = \frac{MSE(K)}{\left\{ \frac{1}{n} \text{tr}[\mathbf{I} - \mathbf{A}(K)] \right\}^2} \quad (2.12)$$



Keterangan:

$$MSE : MSE(K) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_k(x_{ij}))^2$$

$$A(K) : \mathbf{D}(\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T$$

dengan \mathbf{D} merupakan matriks berukuran $n \times (1 + (m \times q) + (m \times r))$.

2.8 Interval Kepercayaan Parameter Regresi

Estimasi interval adalah interval atau jarak yang memiliki probabilitas. Probabilitas ini meliputi nilai batas parameter yang tidak diketahui. Interval ini sering disebut interval kepercayaan dan dapat dinyatakan dalam bentuk persentase (Gujarati, 2015). Menurut Budiantara dkk (2019) jika dimisalkan (X_1, X_2, \dots, X_n) menjadi sampel acak yang saling bebas dengan *probability density function* $(x_i; \theta), \theta \in R$. Misalkan $L(X_1, X_2, \dots, X_n)$ dan $U(X_1, X_2, \dots, X_n)$ menjadi $L(X_1, X_2, \dots, X_n) \leq U(X_1, X_2, \dots, X_n)$.

Interval acak $[L(X_1, X_2, \dots, X_n), U(X_1, X_2, \dots, X_n)]$ adalah interval kepercayaan untuk parameter θ dengan interval kepercayaan dari $1 - \alpha$ ($0 < \alpha < 1$), yang dapat ditulis dalam bentuk Persamaan berikut:

$$P(L(X_1, X_2, \dots, X_n) \leq \theta \leq U(X_1, X_2, \dots, X_n)) = 1 - \alpha \quad (2.10)$$

dengan $L(X_1, X_2, \dots, X_n)$ merupakan batas bawah dari interval kepercayaan, $U(X_1, X_2, \dots, X_n)$ adalah batas atas dari interval kepercayaan dan $1 - \alpha$ merupakan tingkat kepercayaan (Setiawan, 2017).

Terdapat beberapa metode yang dapat digunakan untuk menyusun estimasi interval, salah satunya adalah *pivotal quantity*. *Pivotal quantity* atau pivot merupakan fungsi dari pengamatan dan parameter tidak teramati sehingga distribusi fungsi probabilitas tidak tergantung pada parameter yang tidak diketahui (termasuk parameter *nuisance*). *Pivotal quantity* biasanya digunakan untuk normalisasi agar data dari kumpulan data yang berbeda dapat dibandingkan. *Pivotal quantity* juga menyediakan satu metode untuk membangun interval konfidensi (Shao, 2003). Misalkan $\mathbf{X} = (X_1, X_2, \dots, X_n)$ adalah sebuah sampel *random* dari distribusi yang bergantung pada parameter (atau vektor parameter) θ .

$T(\mathbf{X}, \theta)$ adalah variabel *random* yang distribusinya sama untuk semua θ . $T(\mathbf{X}, \theta)$ disebut sebagai *pivotal quantity* yang dinotasikan sebagai $T(\mathbf{X}, \theta)$ (Setiawan, 2017). Sehingga interval kepercayaan $1 - \alpha$



dengan menggunakan metode *pivotal quantity* adalah (Umam, 2018):

$$P\left(\bar{X} - t_{(n-1; \frac{\alpha}{2})} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{(n-1; \frac{\alpha}{2})} \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha \quad (2.11)$$

2.9 Persentase Penduduk Miskin

Kemiskinan merupakan ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Penduduk dikategorikan sebagai penduduk miskin jika memiliki pengeluaran perkapita per bulan dibawah garis kemiskinan (BPS, 2014). Persentase penduduk yang berada dibawah garis kemiskinan dapat disebut sebagai persentase penduduk miskin. Beberapa faktor yang diduga mempengaruhi persentase penduduk miskin berdasarkan penelitian sebelumnya yang dilakukan oleh (Mariati dkk., 2022), (Dani & Ni'matuzzahroh, 2020) dan (Muhtarom, 2016) sebagai berikut:

1. Angka harapan hidup adalah rata-rata seseorang diharapkan hidup dalam suatu tempat atau kelompok orang.
2. Rata-rata Lama Sekolah adalah rata-rata jumlah tahun yang dihabiskan oleh penduduk berusia 15 tahun keatas untuk menempuh semua jenis pendidikan yang pernah dijalani.
3. 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 dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut.

