

**PEMODELAN REGRESI SEMIPARAMETRIK
SMOOTHING SPLINE PADA DATA LONGITUDINAL
(Studi Kasus Kemiskinan di Provinsi Sulawesi Tenggara)**

SKRIPSI



SELI LISNAYATI

H051191033

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

MAKASSAR

2024

**PEMODELAN REGRESI SEMIPARAMETRIK
SMOOTHING SPLINE PADA DATA LONGITUDINAL
(Studi Kasus Kemiskinan di Provinsi Sulawesi Tenggara)**

SKRIPSI

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

**SELI LISNAYATI
H051191033**

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

**MAKASSAR
FEBRUARI 2024**

LEMBAR PERNYATAAN KEOTENTIKAN

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

**PEMODELAN REGRESI SEMIPARAMETRIK *SMOOTHING SPLINE*
PADA DATA LONGITUDINAL
(Studi Kasus Kemiskinan di Provinsi Sulawesi Tenggara)**

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

Makassar, 16 Februari 2024



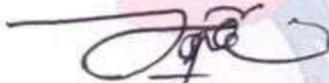
Seli Lisnawati
NIM H051191033

**PEMODELAN REGRESI SEMIPARAMETRIK
SMOOTHING SPLINE PADA DATA LONGITUDINAL
(Studi Kasus Kemiskinan di Provinsi Sulawesi Tenggara)**

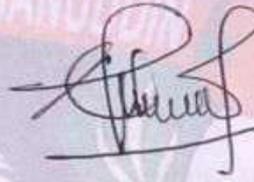
Disetujui Oleh:

Pembimbing Utama

Pembimbing Pertama



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



Dra. Nasrah Sirajang, M.Si.
NIP. 196505191993032002

Ketua Program Studi



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

Pada 16 Februari 2024

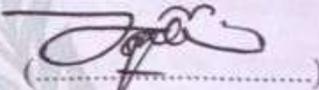
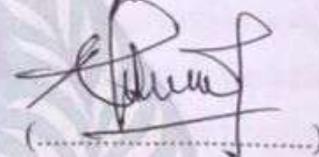
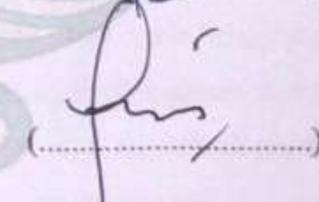
HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Seli Lisnayati
NIM : H051191033
Program Studi : Statistika
Judul Skripsi : *Pemodelan Regresi Semiparametrik Smoothing Spline*
Pada Data Longitudinal (Studi Kasus Kemiskinan di
Provinsi Sulawesi Tenggara)

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.

UNIVERSITAS HASANUDDIN
DEWAN PENGUJI

1. Ketua : Dr. Anna Islamiyati, S.Si., M.Si. 
2. Sekretaris : Dra. Nasrah Sirajang, M.Si. 
3. Anggota : Prof. Dr. Nurtiti Sunusi, S.Si., M.Si. 
4. Anggota : Dr. Nirwan, M.Si 

Ditetapkan di : Makassar

Tanggal : 16 Februari 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 “**Pemodelan Regresi Semiparametrik *Smoothing Spline* pada Data Longitudinal (Studi Kasus Kemiskinan di Provinsi Sulawesi Tenggara)**” 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 **La Zaini** dan Ibunda **Samria** 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-kakak tersayang penulis **Yoyon Sularso** dan **Hendriyono** serta adik tersayang penulis **Elin Damayati** 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 dan ketulusan juga penulis ucapkan kepada:

1. **Bapak 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. **Ibu Dra. Nasrah Sirajang, 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. **Ibu Prof. Dr. Nurtiti Sunusi, 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 Dr. Nirwan, 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 *Pondok Putri Innawa* **Aulia Karimah, Izzatin Rumaisha Zahra, Kania Meliani Kaharuddin**, dan **Waode Andini Putri Sukma Laudia**. Terima kasih atas kebersamaan, kebahagiaan, kekeluargaan, bantuan serta kebaikannya terhadap penulis sejak menjadi mahasiswa baru dan tetap menemani penulis hingga menyelesaikan masa studi.
8. Teman-teman **Andi Muhammad Rajab, Muhammad Syamsul Bahri, Muliana, Nurul Hikmah, Nur Aisyah** dan **Vinaya Rifqi Anandari**. Terima

kasih atas kebersamaan, kebahagiaan, bantuan, serta kebaikannya terhadap penulis.

9. Teman-teman **KKN Gelombang 108 Kecamatan Tompobulu Desa Bontomanai**.
10. Teman lain **Inda Mutmainna, Rahmah Ningsih D.J, Mufa, dan Jule**. 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, 16 Februari 2024

Seli Lisnayati

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

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

Nama : Seli Lisnayati
NIM : H051191033
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:

“Pemodelan Regresi Semiparametrik *Smoothing Spline* pada Data Longitudinal (Studi Kasus Kemiskinan di Provinsi Sulawesi Tenggara)”

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 16 Februari 2024.

Yang menyatakan,

(Seli Lisnayati)

ABSTRAK

Model regresi umumnya dibagi ke dalam tiga bentuk, yaitu model regresi parametrik, model regresi nonparametrik, dan model regresi semiparametrik. Regresi semiparametrik digunakan apabila sebagian bentuk kurva regresi diasumsikan diketahui dan sebagian lagi tidak diketahui bentuk kurva regresinya yang merupakan kombinasi regresi parametrik dan regresi nonparametrik. Kemampuan untuk menemukan bentuk pola kurva regresi yang belum diketahui didukung dengan adanya parameter pada setiap jenis metode regresi nonparametrik yang mengestimasi kurva regresi menjadi lebih fleksibel. *Spline* merupakan model yang sangat baik dalam memodelkan data yang memiliki pola yang berubah-ubah pada sub-sub interval tertentu dan memiliki fleksibilitas yang tinggi. Dari beberapa estimator *spline*, *smoothing spline* merupakan fungsi yang mampu memetakan data dengan baik dan mempunyai variansi *error* yang kecil. Penelitian ini bertujuan untuk mendapatkan bentuk estimasi regresi semiparametrik *smoothing spline* menggunakan data longitudinal pada kasus kemiskinan di Provinsi Sulawesi Tenggara. Hasil analisis yang diperoleh, didapatkan estimasi $\hat{y} = (\mathbf{A}_{par} + \mathbf{A}_{nonpar})\mathbf{y}$. Adapun model semiparametrik *smoothing spline* terbaik dari kasus kemiskinan berdasarkan nilai GCV minimum yaitu 0,464 dengan nilai $R^2 = 95,98\%$. Hal tersebut bermakna bahwa variabel tingkat pengangguran terbuka, gini ratio, indeks pembangunan manusia dan harapan lama sekolah dapat menjelaskan variabel persentase penduduk miskin di Provinsi Sulawesi Tenggara sebesar 95,98% sedangkan sisanya dijelaskan oleh variabel lain yang tidak digunakan dalam penelitian ini.

Kata Kunci: Data Longitudinal, Kemiskinan, Regresi Semiparametrik, *Smoothing Spline*.

ABSTRACT

Regression models are generally divided into three forms, namely parametric regression models, nonparametric regression models, and semiparametric regression models. Semiparametric regression is used when part of the shape of the regression curve is assumed to be known and part of it is unknown, which is a combination of parametric regression and nonparametric regression. The ability to find the unknown shape of the regression curve pattern is supported by the existence of parameters in each type of nonparametric regression method that estimates the regression curve to be more flexible. Spline is an excellent model in modeling data that has a changing pattern at certain sub-intervals and has high flexibility. From several spline estimators, smoothing spline is a function that is able to map data well and has a small error variance. This study aims to obtain the form of smoothing spline semiparametric regression estimation using longitudinal data on poverty cases in Southeast Sulawesi Province. The results of the analysis obtained, the estimate $\hat{\mathbf{y}} = (\mathbf{A}_{\text{par}} + \mathbf{A}_{\text{nonpar}})\mathbf{y}$ is obtained. The best semiparametric smoothing spline model of the poverty case based on the minimum GCV value is 0,464 with a value of $R^2 = 95,98\%$. This means that the variables of open unemployment rate, gini ratio, human development index and expected years of schooling can explain the variable percentage of poor people in Southeast Sulawesi Province by 95,98% while the rest is explained by other variables not used in this research.

Keywords: *Longitudinal Data, Poverty, Semiparametric Regression, Spline Smoothing.*

DAFTAR ISI

HALAMAN JUDUL	ii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSETUJUAN PEMBIMBING	iv
HALAMAN PENGESAHAN	v
KATA PENGANTAR.....	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	ix
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Data Longitudinal.....	4
2.2 Analisis Regresi Parametrik.....	4
2.3 Analisis Regresi Nonparametrik	5
2.4 Analisis Regresi Semiparametrik.....	6
2.5 Estimator <i>Smoothing Spline</i> pada Data Longitudinal	6
2.6 Pemilihan Parameter Penghalus Optimal.....	8
2.7 Kemiskinan	8
BAB III METODOLOGI PENELITIAN	11
3.1 Data	11
3.2 Metode Analisis	11
BAB IV HASIL DAN PEMBAHASAN.....	13
4.1 Estimasi Parameter Model Regresi Semiparametrik <i>Smoothing Spline</i>	13

4.2	Memodelkan Data Persentase Penduduk Miskin Menggunakan Regresi Semiparametrik <i>Smoothing Spline</i>	19
BAB V KESIMPULAN DAN SARAN		31
5.1	Kesimpulan	31
5.2	Saran.....	31
DAFTAR PUSTAKA		33
LAMPIRAN.....		36

DAFTAR TABEL

Tabel 2. 1 Struktur Data Longitudinal.....	4
Tabel 3. 1 Variabel Respon dan Prediktor.....	11
Tabel 4. 1 Statistika Deskriptif.....	19
Tabel 4. 2 Nilai GCV Minimum.....	22
Tabel 4. 3 Nilai Parameter $\hat{\beta}$	23
Tabel 4. 4 Nilai Parameter \hat{a}_i	24
Tabel 4. 5 Nilai Parameter \hat{c}_{ij}	25

DAFTAR GAMBAR

Gambar 4. 1 Scatter Plot PPM dan TPT	20
Gambar 4. 2 Scatter Plot PPM dan GR	21
Gambar 4. 3 Scatter Plot PPM dan IPM.....	21
Gambar 4. 4 Scatter Plot PPM dan HLS	22

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan suatu analisis yang membahas tentang ketergantungan antara variabel respon dan variabel prediktor. Terdapat dua tujuan utama dalam analisis regresi yaitu mengeksplorasi dan memprediksi hubungan antara variabel respon dan variabel prediktor (Rosadi dkk., 2022). Model regresi umumnya dibagi ke dalam tiga bentuk, yaitu model regresi parametrik, model regresi nonparametrik, dan model regresi semiparametrik. Bentuk model regresi tergantung pada kurva regresinya (Hidayat dkk., 2017). Apabila bentuk kurva regresi diketahui misalnya linier, kuadratik, kubik, polinomial derajat p , eksponen dan lainnya maka pendekatan model regresi tersebut merupakan model regresi parametrik. Apabila bentuk kurva regresi tidak diketahui pola bentuknya maka model regresi tersebut merupakan model regresi nonparametrik (Dani dkk., 2021), sedangkan model regresi semiparametrik merupakan kombinasi antara regresi parametrik dan nonparametrik. Kombinasi ini dimaksudkan bahwa dalam analisis regresi semiparametrik memuat sekaligus model regresi parametrik dan model regresi nonparametrik (Wibowo dkk., 2013).

Regresi semiparametrik digunakan apabila sebagian bentuk kurva regresi diasumsikan diketahui dan sebagian lagi tidak diketahui bentuk kurva regresinya. Kemampuan untuk menemukan bentuk pola kurva regresi yang belum diketahui didukung dengan adanya parameter pada setiap jenis metode regresi nonparametrik yang mengestimasi kurva regresi menjadi lebih fleksibel. Ada beberapa teknik estimasi dalam regresi nonparametrik antara lain estimator histogram, *Spline*, kernel, polinomial lokal, deret fourier, dan *wavelet* (Adrianingsih dkk., 2020). Diantara beberapa estimator tersebut, yang sering digunakan dalam mengestimasi kurva regresi yaitu estimator *spline*.

Spline merupakan model yang sangat baik dalam memodelkan data yang memiliki pola yang berubah-ubah pada sub-sub interval tertentu dan memiliki fleksibilitas yang tinggi (Ningrum dkk., 2020). *Spline* adalah potongan polinomial orde p dengan titik bersama, dari potongan-potongan tersebut disebut dengan knot. Titik knot merupakan perpaduan dua kurva yang menunjukkan pola

perubahan perilaku kurva pada selang yang berbeda. Penggunaan titik knot banyak digunakan dalam regresi nonparametrik, karena secara visual dapat menunjukkan setiap perubahan pola perilaku yang terjadi dalam interval waktu tertentu (Islamiyati, 2014).

Beberapa pengembangan estimator *spline* antara lain *truncated spline*, *smoothing spline*, *penalized spline*, *robust spline*, *b-spline*, *multiresponse spline*, dan lain-lain (Islamiyati dkk., 2022). Dari beberapa estimator *spline*, *smoothing spline* merupakan fungsi yang mampu memetakan data dengan baik dan mempunyai variansi error yang kecil. Estimasi *smoothing spline* berdasarkan kriteria model *smoothing* kurva dan ukuran yang telah ditentukan oleh parameter *smoothing*. Secara umum *smoothing spline* menggunakan pendekatan *Penalized Least Square* (PLS). Namun, jika terdapat autokorelasi maka PLS tidak dapat digunakan sehingga alternatif yang dapat mengakomodir autokorelasi adalah *Penalized Weighted Least Square* (PWLS) (A. A. Fernandes dkk., 2018). Beberapa penelitian yang telah menggunakan estimasi *smoothing spline* di antaranya, yaitu Ghafar (2017) menerapkan analisis regresi semiparametrik *smoothing spline* dengan menggunakan penalti dan tanpa penalti pada data tingkat kemiskinan di Jawa Timur, Chofifi (2018) menduga model regresi semiparametrik *smoothing spline* dengan pendekatan PWLS, dan Risnawati dkk. (2019) menggunakan estimator *smoothing spline* pada data longitudinal. Estimasi *smoothing spline* telah digunakan dalam berbagai penelitian, termasuk pada data longitudinal.

Data longitudinal merupakan gabungan dari data *cross-section* dan *time-series* yang diamati dan diukur berulang kali pada suatu interval waktu tertentu dalam subyek yang sama saling berkorelasi (Zia dkk., 2017). Keuntungan menggunakan data longitudinal yaitu memberikan data yang lebih informatif, mengurangi kolinearitas antar variabel, menguji dan membangun model yang lebih kompleks dibandingkan dengan menggunakan data *cross-section* atau *time-series* (Ratnasari dkk., 2014). Berdasarkan uraian tersebut, peneliti akan menganalisis data longitudinal dengan regresi semiparametrik *smoothing spline* dan diaplikasikan pada kasus kemiskinan di Provinsi Sulawesi Tenggara dari tahun 2018 – 2022.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana bentuk estimasi parameter model regresi semiparametrik *smoothing spline* dengan data longitudinal pada kasus kemiskinan di Sulawesi Tenggara?
2. Bagaimana model kasus kemiskinan di Provinsi Sulawesi Tenggara menggunakan regresi semiparametrik *smoothing spline* pada data longitudinal?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan yaitu data longitudinal berupa data *time series* dari tahun 2018-2022 dan data *cross section* meliputi 17 kabupaten/kota di Sulawesi Tenggara.
2. Pemilihan model pada analisis semiparametrik *smoothing spline* pada data longitudinal dengan pendekatan PWLS.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter model regresi semiparametrik *smoothing spline* dengan data longitudinal pada kemiskinan di Sulawesi Tenggara.
2. Memperoleh model kasus kemiskinan menggunakan regresi semiparametrik *smoothing spline* pada data longitudinal.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Memperoleh wawasan baru mengenai model analisis regresi semiparametrik *smoothing spline* pada data longitudinal.
2. Menjadi bahan bacaan atau informasi bagi pembaca mengenai pembentukan model analisis regresi semiparametrik *smoothing spline* pada data longitudinal sehingga dapat dikembangkan lebih lanjut menggunakan analisis ini.

BAB II

TINJAUAN PUSTAKA

2.1 Data Longitudinal

Data longitudinal adalah data yang diamati dan diukur berulang kali pada suatu interval waktu tertentu, dengan kata lain data longitudinal merupakan gabungan antara data silang (*cross section*) dengan data runtun waktu (*time series*). Unit *cross section* berupa individu, perusahaan dan wilayah atau lokasi, sedangkan unit *time series* berupa harian, bulanan dan tahunan (Amaliah dkk., 2020). Data longitudinal memiliki keunggulan yaitu memiliki kemampuan untuk mengenali pengaruh waktu terhadap respon. Data longitudinal diasumsikan obyek-obyek saling independen satu sama lainnya, tetapi antara pengamatan di dalam obyek yang sama saling dependen, karena cenderung berkorelasi (Islamiyati, 2014). Struktur data longitudinal dapat ditunjukkan pada Tabel 1.

Tabel 2. 1 Struktur Data Longitudinal

Subyek	Waktu	Variabel	Variabel	Variabel	...	Variabel
(i)	(j)	(Y_{ij})	(X_{ij1})	(X_{ij2})		(X_{ijq})
Subyek 1	1	Y_{11}	X_{111}	X_{112}	...	X_{11q}
	2	Y_{12}	X_{121}	X_{122}	...	X_{12q}
	\vdots	\vdots	\vdots	\vdots	...	\vdots
	j_1	Y_{1j_1}	X_{1j_11}	X_{1j_12}	...	X_{1j_1q}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Subyek n	1	Y_{n1}	X_{n11}	X_{n12}	...	X_{n1q}
	2	Y_{n2}	X_{n21}	X_{n22}	...	X_{n2q}
	\vdots	\vdots	\vdots	\vdots	...	\vdots
	j_n	Y_{nt_n}	X_{nj_n1}	X_{nj_n2}	...	X_{nj_nq}

2.2 Analisis Regresi Parametrik

Analisis regresi memiliki beberapa pendekatan untuk menduga model regresi parametrik dan regresi nonparametrik. Budiantara (2006) menyebutkan bahwa regresi parametrik digunakan untuk mengetahui pola hubungan antara variabel respon dengan variabel prediktor yang diasumsikan telah diketahui bentuk fungsinya, sedangkan bila pola hubungannya tidak diketahui maka pendekatan

nonparametrik lebih sesuai untuk menyelesaikan masalah jika data tidak linier. Salah satu bentuk regresi parametrik dapat dinyatakan sebagai model regresi linier berganda sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \quad (2.1)$$

Dalam upaya mencari persamaan garis yang paling tepat (yang menghasilkan galat yang kecil), pendugaan parameter regresi dapat dibayangkan sebagai upaya memilih model yang membuat jumlah kuadrat galat terhadap pengamatan menjadi sekecil-kecilnya. Salah satu metode yang dapat digunakan untuk menduga parameter adalah *Ordinary Least Square* (OLS) (Andreas dkk., 2021). Menurut Weisberg (2005) persamaan (2.4) dapat dituliskan dalam bentuk matriks berikut:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

Keterangan:

y_i : vektor dari variabel respon

x_{ij} : nilai variabel prediktor ke- i dari amatan ke- j

$\boldsymbol{\beta}$: vektor parameter

\mathbf{X} : matriks dari variabel prediktor

2.3 Analisis Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan dalam analisis regresi yang digunakan apabila kurva regresinya tidak diasumsikan memiliki bentuk tertentu. Dalam regresi nonparametrik, kurva regresi hanya diasumsikan halus (*smooth*), sehingga pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi karena data diharapkan mencari sendiri bentuk estimasi kurva regresi tanpa dipengaruhi oleh faktor subjektivitas peneliti. Secara umum, model regresi nonparametrik adalah (Sholikha dkk., 2019):

$$y_i = f(t_i) + \varepsilon_i, i = 1, 2, \dots, N \quad (2.3)$$

y_i menyatakan variable respon pada subjek ke- i , $f(t_i)$ menyatakan variabel yang tidak diketahui kurva atau polanya, t_i menyatakan variabel prediktor pada subjek ke- i , dan ε_i merupakan *error* pengukuran pada subjek ke- i . Model regresi nonparametrik pada data longitudinal, dikembangkan dengan pengamatan N subjek yang diamati secara berulang dalam J waktu dapat dituliskan sebagai berikut (Fernandes dkk., 2014):

$$y_{ij} = f(t_{ijq}) + \varepsilon_{ij} \quad (2.4)$$

dengan:

$$i = 1, 2, \dots, N \quad j = 1, 2, \dots, J \quad q = 1, 2, \dots, Q$$

Keterangan:

y_{ij} : respon pada subjek ke- i dan pengamatan pada waktu ke- j .

t_{ij} : prediktor pada subjek ke- i dan pengamatan pada waktu ke- j .

f_i : kurva regresi hubungan prediktor dengan respon pada subjek ke- i .

N : banyaknya subjek.

J : banyaknya pengamatan setiap subjek.

ε_{ij} : *random error* pada subjek ke- i dan pengamatan pada waktu ke- j .

2.4 Analisis Regresi Semiparametrik

Dalam analisis regresi tidak semua peubah penjelas dapat didekati dengan pendekatan parametrik, karena tidak adanya informasi tentang bentuk hubungan antara variabel prediktor dan variabel responnya, sehingga harus dilakukan pendekatan nonparametrik. Menurut Purnomo dkk. (2008), dengan menggabungkan pendekatan parametrik dan nonparametrik dalam suatu pendekatan regresi akan didapatkan suatu model semiparametrik. Persamaan semiparametrik linier sederhana dituliskan sebagai berikut.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + f(t_{ij}) + \varepsilon_{ij} \quad (2.5)$$

Keterangan:

y_{ij} : respon pada subjek ke- i dan pengamatan pada waktu ke- j .

x_{ij} : prediktor pada subjek ke- i dan pengamatan pada waktu ke- j untuk komponen parametrik

t_{ij} : prediktor pada subjek ke- i dan pengamatan pada waktu ke- j untuk komponen non-parametrik

$f(t_{ij})$: fungsi nonparametrik yang tidak diketahui bentuknya

ε_{ij} : galat yang memiliki rata-rata 0 dan ragam σ^2

2.5 Estimator *Smoothing Spline* pada Data Longitudinal

Spline merupakan salah satu pendekatan dalam regresi semiparametrik. Digunakan pendekatan *spline* karena mampu menjelaskan perubahan perilaku

fungsi dalam sub-interval tertentu dan dapat mengatasi atau mengurangi pola data yang mengalami peningkatan tajam serta memiliki sifat yang fleksibel (Kosmaryati & Kartikasari, 2020). Model regresi nonparametrik pada Persamaan (2.5) tidak memberikan asumsi terhadap bentuk kurva regresi f . Kurva regresi f diasumsikan *smooth*, dalam arti termuat dalam ruang Sobolev $W_2^m[a, b]$. Dalam regresi nonparametrik, kurva regresi diasumsikan *smooth* dalam arti kontinu dan terdiferensiasi. Asumsi ini penting agar diperoleh kurva hampiran yang mulus dan bukan merupakan kurva dengan perubahan cepat. Jika ada asumsi demikian, maka harus dicari suatu ruang fungsi yang sesuai. Salah satu ruang yang memenuhi adalah ruang Sobolev:

$$W_2^m([a, b]) = \left\{ \int_b^a (f^{(m)}(t))^2 dt < \infty \right\} \quad (2.6)$$

Ruang Sobolev menjamin bahwa fungsi tersebut memiliki turunan. Hal ini sesuai dengan teorema yang menyatakan bahwa jika f memiliki turunan pada $t = c$, maka f kontinu pada $t = c$. Secara matematis, ada banyak kriteria pemulusan kurva. Kriteria yang paling mendasar adalah kontinuitas, dan gagasan yang lebih tegas lagi adalah fungsi yang diturunkan.

Pendugaan parameter regresi nonparametrik dapat diselesaikan dengan meminimumkan optimasi PWLS sebagai berikut (Risnawati dkk., 2019):

$$\text{Min} \left\{ M^{-1}(\mathbf{y} - \mathbf{f})^T \mathbf{W}(\mathbf{y} - \mathbf{f}) + \sum_{i=1}^N \lambda_i \int_{a_i}^{b_i} (f_i^{(m)}(t_{ij}))^2 dt_{ij} \right\} \quad (2.7)$$

dengan $M = NJ$, $\mathbf{y} = (y_{11}, y_{12}, \dots, y_{NJ})^T$, $\mathbf{f} = (f_1, f_2, \dots, f_N)$ dan $f_i = (f_1(t_{11}), f_2(t_{12}), \dots, f_N(t_{NJ}))^T$. Matriks pembobot dalam Persamaan (2.7) yaitu:

$$\mathbf{W}^{-1} = \text{diag}(\mathbf{W}_1^{-1}, \mathbf{W}_2^{-1}, \dots, \mathbf{W}_N^{-1}) \text{ dengan } \mathbf{W}^{-1} = \frac{(y_i - f_i)(y_i - f_i)^T}{J}; i = 1, 2, \dots, N.$$

Nilai \hat{f}_i diperoleh dengan optimasi PLS atau OLS.

Optimasi PWLS pada Persamaan (2.7) dipertimbangkan parameter penghalus λ_i sebagai pengontrol antara *goodness of fit* (ruas pertama) dan *roughness penalty* (ruas kedua). Penguraian optimasi PWLS di atas menjadi *goodness of fit* (Risnawati dkk., 2019).

$$R(f) = M^{-1}(\mathbf{y} - \mathbf{f})^T \mathbf{W}(\mathbf{y} - \mathbf{f}) \quad (2.8)$$

dan *roughness penalty*

$$J(f) = \sum_{i=1}^N \lambda_i \int_{a_i}^{b_i} (f_i^{(m)}(t_{it}))^2 dt_{it} \quad (2.9)$$

Keterangan:

- M : banyaknya jumlah observasi
- λ_i : parameter penghalus
- \mathbf{W} : matriks pembobot (varian-kovarian)

2.6 Pemilihan Parameter Penghalus Optimal

Pemilihan parameter yang optimal menjadi salah satu hal yang penting, karena berkaitan dengan perolehan estimator *spline* yang optimal (Howell, 2007). Jika nilai parameter pemulus sangat kecil (λ mendekati nol) maka penduga *spline* akan sangat kasar, sedangkan jika parameter pemulus sangat besar (λ mendekati tak terhingga) maka penduga *spline* akan sangat mulus (Eubank, 1999). Salah satu metode yang banyak digunakan untuk memilih parameter penghalus adalah *Generalized Cross Validation* (GCV). Menurut (Risnawati dkk., 2019), pemilihan parameter pemulus yang optimal berdasarkan pendekatan PWLS pada persamaan berikut:

$$GCV = \frac{MSE(\lambda)}{\left(\frac{1}{M} \text{trace}(\mathbf{I} - \mathbf{A}_\lambda)\right)^2} = \frac{\frac{1}{M} \mathbf{y}^T (\mathbf{I} - \mathbf{A}_\lambda)^T (\mathbf{I} - \mathbf{A}_\lambda) \mathbf{y}}{\left(\frac{1}{M} \text{trace}(\mathbf{I} - \mathbf{A}_\lambda)\right)^2} \quad (2.10)$$

dengan $MSE(\lambda) = \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^J (y_{ij} - f(t_{ij}))^2$; $f(t) = \mathbf{A}_\lambda \mathbf{y}$

2.7 Kemiskinan

Kemiskinan adalah suatu keadaan yang menyangkut ketidakmampuan dalam memenuhi tuntutan kehidupan yang paling minimum, khususnya dari aspek konsumsi dan pendapatan (Jacobus dkk., 2018). Kemiskinan merupakan masalah global, yang sering berkaitan dengan kebutuhan hidup, kesulitan hidup serta kekurangan di semua aspek kehidupan. Masalah kemiskinan memang telah lama ada sejak dahulu, di mana banyak negara yang sedang berkembang mengalami laju pertumbuhan ekonomi yang tinggi tetapi tidak membawa manfaat yang berguna bagi penduduknya. Ini dikarenakan tingkat kesenjangan yang terjadi juga semakin

besar dan jumlah kemiskinan tetap saja banyak. Indonesia juga mengalami masalah tersebut, dari sejak masa kemerdekaan hingga saat ini (Priyadi & Asmoro, 2011).

Persentase penduduk miskin merupakan persentase penduduk yang berada dibawah garis kemiskinan. Adapun persentase penduduk miskin secara matematis menurut Badan Pusat Statistik (BPS) diperoleh dari:

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^{\alpha}$$

dengan, $\alpha = 0$, z merupakan garis kemiskinan, y_i merupakan rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ($i = 1, 2, \dots, q$) dengan $y_i < z$, sedangkan q merupakan banyaknya penduduk yang berada di bawah garis kemiskinan serta n merupakan jumlah penduduk.

Berdasarkan BPS persentase penduduk miskin di Indonesia meningkat dari tahun 2018 sebesar 9,66% hingga 2022 sebesar 9,71%. Pada tahun 2022 data dari BPS, Provinsi Sulawesi Tenggara menempati urutan ke 20 dari 34 provinsi dengan persentase penduduk miskin terbanyak di Indonesia.

Problematika kemiskinan yang terjadi di kabupaten/kota di Provinsi Sulawesi Tenggara terutama tidak meratanya persentase penduduk miskin, oleh karena itu perlunya diketahui faktor-faktor yang berperan dalam peningkatan persentase kemiskinan. Beberapa faktor yang diduga mempengaruhi persentase penduduk miskin berdasarkan penelitian sebelumnya yaitu sebagai berikut:

1. Tingkat pengangguran terbuka memiliki hubungan yang sangat erat dalam mempengaruhi tingkat kemiskinan. Pengangguran yang tinggi akan menyebabkan pendapatan berkurang sehingga tidak dapat memenuhi kebutuhan sehari-hari yang pada akhirnya akan mengalami kemiskinan (Ibrahim, 2021). Menurut BPS, secara matematis Tingkat pengangguran terbuka diperoleh dari,

$$\frac{\text{jumlah pengangguran} \times 100\%}{\text{jumlah tingkatan kerja}}$$

2. Gini ratio menggambarkan pemerataan dan ketimpangan secara keseluruhan, mulai dari pendapatan hingga distribusi. Koefisien gini berkisar antara 0 sampai 1, dengan interpretasi semakin mendekati angka 0 semakin baik dalam distribusi pendapatannya, sedangkan jika koefisien gini mendekati 1 maka

tingkat ketimpangan atau ketidakmerataan semakin tinggi (Islami & Fitrianto, 2023).

3. Indeks pembangunan manusia menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya serta menjadi indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk) (Priseptian & Primandhana, 2022).
4. Harapan lama sekolah didefinisikan sebagai lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Diasumsikan bahwa peluang anak tersebut akan tetap bersekolah pada umur-umur berikutnya sama dengan peluang penduduk yang bersekolah per jumlah penduduk untuk umur yang sama saat ini. Menurut BPS, secara matematis harapan lama sekolah dapat diperoleh dari,

$$HLS_{\alpha}^t = FK \times \sum_{i=\alpha}^k \frac{E_i^t}{P_i^t}$$

dengan HLS_{α}^t adalah harapan lama sekolah pada umur α di tahun t , FK adalah faktor koreksi pesantren, E_i^t jumlah penduduk usia i yang bersekolah pada tahun t , P_i^t adalah penduduk usia i pada tahun t dan i adalah usia $(\alpha, \alpha + 1, \dots, n)$.