

SKRIPSI

**PENGUJIAN HIPOTESIS MODEL REGRESI NONPARAMETRIK SPLINE
TRUNCATED DALAM PEMODELAN TINGKAT PENGANGGURAN
TERBUKA DI SULAWESI SELATAN**

Disusun dan diajukan oleh

MAR'ATUL WILDANI SUDARMIN

H12116003



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

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

**PENGUJIAN HIPOTESIS MODEL REGRESI NONPARAMETRIK SPLINE
TRUNCATED DALAM PEMODELAN TINGKAT PENGANGGURAN
TERBUKA DI SULAWESI SELATAN**

SKRIPSI

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

MAR'ATUL WILDANI SUDARMIN

H 121 16 003

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

MAKASSAR

2021

LEMBAR PENGESAHAN (TUGAS AKHIR)

PENGUJIAN HIPOTESIS MODEL REGRESI NONPARAMETRIK SPLINE
TRUNCATED DALAM PEMODELAN TINGKAT PENGANGGURAN
TERBUKA DI SULAWESI SELATAN

Disusun dan diajukan oleh

MAR'ATUL WILDANI SUDARMIN
H 121 16 003

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal April 2021
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Dr. Dr. Georgina Maria Tinungki, M.Si
NIP. 19620926 198702 2001

Dra. Nasrah Sirajang, M.Si
NIP. 19650519 199303 2002

Ketua Program Studi,




Dr. Nurtiti Sunusi, S.Si, M.Si
NIP. 19720117 199703 2002

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Mar'atul Wildani Sudarmin
NIM : H 121 16 003
Program Studi : Statistika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul:

Pengujian Hipotesis Model Regresi Nonparametrik Spline Truncated dalam
Pemodelan Tingkat Pengangguran Terbuka Di Sulawesi Selatan

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 19 April 2021

Yang menyatakan,



Mar'atul Wildani Sudarmin

KATA PENGANTAR

Alhamdulillahirobbil'alamin. Segala puji penulis haturkan atas kehadiran Allah SWT, karena atas rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul “Pengujian Hipotesis Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Tingkat Pengangguran Terbuka Di Sulawesi Selatan” ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains (S.Si) pada Pogram Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Univesitas Hasanuddin. Shalawat serta salam semoga tetap tercurahkan kepada junjungan kita Nabi Muhammad SAW karena Beliaulah yang membawa umat manusia dari kegelapan menuju ke alam yang terang benderang.

Penulis menyadari bahwa dalam penulisan skripsi ini tidak terlepas dari dorongan, dukungan, bimbingan, serta kerjasama dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penulisan skripsi ini. Terutama kepada orang tua, **Ayahanda Sudarmin** dan **Ibunda Fatmawati Zainuddin** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran, limpahan cinta dan kasih sayang tiada tara, serta telah memberikan dukungan dan doa kepada penulis dalam menyelesaikan pedidikan. Kepada kedua adikku **Muh. Fahmi Sayyid Al Maulana** dan **Muh. Said Al Gifari** yang selalu sabar dan memberikan dukungan penuh kepada penulis. Serta untuk keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini.

Demikian pula dengan penuh keikhlasan dan kerendahan hati penulis mengucapkan penghargaan dan terima kasih kepada :

1. **Ibu Prof. Dr. Dwia Aries Tina Palubuhu, MA**, 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. Nurtiti Sunusi, S.Si, M.Si**, selaku Ketua Departemen Statistika serta segenap dosen pengajar dan staf Departemen Statistika yang telah membekali

ilmu dan memberikan kemudahan-kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.

4. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si**, selaku Pembimbing Utama dan **Ibu Dra. Nasrah Sirajang, M.Si** selaku Pembimbing Pertama sekaligus Penasehat Akademik yang telah sabar dan ikhlas meluangkan begitu banyak waktu dan pemikirannya untuk membimbing dan memberikan masukan dalam penulisan skripsi ini.
5. **Ibu Anisa, S.Si, M.Si** dan **Ibu Dr. Anna Islamiyati, S.Si, M.Si** selaku tim penguji atas semua saran dan kritikan yang membangun dalam penyempurnaan penyusunan skripsi ini serta waktu yang telah diberikan kepada penulis.
6. Seluruh **Dosen dan Staf Departemen Statistika** yang senantiasa berbagi ilmu, nasehat, dan motivasi selama menjadi mahasiswa di Departemen Statistika.
7. Sahabat seperjuangan **Statistika 2016** terkhusus **Tim Marketing (Iis, Hajriah, dan Risma), Isna, Imma, Inchi, Ayu Riski, Jumri, Atiek, Cacong**, dan yang banyak membantu penulis selama perkuliahan. Untuk **Tim Marketing**, terimakasih pula selalu bersedia menjadi tempat curhat penulis terlebih dalam proses penyelesaian skripsi ini.
8. Keluarga **Acceleration 05** terkhusus **Astuti, Shinta, Tiong**, dan **Darmi** yang telah memberikan banyak dukungan dan motivasi kepada penulis. Penulis sangat bersyukur karena telah dipertemukan dengan orang-orang hebat seperti kalian.
9. Teman-teman **KKN Tematik UNHAS Desa Sehat Gowa Gel. 102** terkhusus posko Bissoloro. Terima kasih telah menjadi teman sekaligus keluarga selama sebulan lebih, semoga silaturahmi kita tetap terjalin.
10. Teman-teman **SMAN 1 Sungguminasa** terkhusus **Risya, Ayu, Nurul**, dan **Nisa** yang selalu memberikan dukungan dan semangat kepada penulis.
11. Terimakasih kepada **Pinkers** terkhusus **Vee, Hana, Windy, Qurr, Nu, Kiki, Chaca**, dan **Kia** yang selalu memberikan semangat kepada penulis.

12. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah Subhanahu Wata'ala.

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

Wassalamualaikum Warahmatullahi Wabarakatuh.

Makassar, 19 April 2021

Penulis

ABSTRAK

Sulawesi Selatan menjadi salah satu daerah penyumbang penganggur tertinggi di Indonesia. Pada Februari 2020 tercatat angka pengangguran di Sulawesi Selatan sebesar 6,07 persen menjadikan Sulawesi Selatan terdata masuk lima besar setelah Banten, Jawa Barat, Maluku, Kalimantan Timur dan Papua Barat sebagai daerah dengan pengangguran tertinggi. Maka dilakukan penelitian untuk menganalisis faktor yang diduga menjadi penyebab terjadinya tingkat pengangguran terbuka (TPT) di Sulawesi Selatan tahun 2020. Tujuan penelitian ini adalah menganalisis faktor-faktor yang diduga mempengaruhi TPT di Sulawesi Selatan. Berdasarkan *scatterplot* diperoleh bahwa tidak terdapat pola tertentu antara TPT dengan faktor-faktor yang diduga berpengaruh sehingga digunakan metode regresi nonparametrik spline *truncated*. Hasil analisis menunjukkan model regresi nonparametrik spline *truncated* terbaik dalam memodelkan TPT di Sulawesi Selatan adalah dengan tiga titik knot. Model yang dihasilkan memiliki nilai GCV yang paling minimum sebesar 2,16 dengan R^2 sebesar 98,21%. Berdasarkan hasil pengujian secara simultan dan parsial didapatkan bahwa persentase penduduk miskin, rata-rata lama sekolah, *dependency ratio*, dan indeks pembangunan manusia berpengaruh secara signifikan terhadap tingkat pengangguran terbuka di Sulawesi Selatan tahun 2020.

Kata Kunci : Regresi Nonparametrik, Spline *Truncated*, Tingkat Pengangguran Terbuka (TPT)

ABSTRACT

South Sulawesi is one of the provinces with the highest rates of unemployment in Indonesia. In February 2020, South Sulawesi's unemployment rate was recorded at 6.07 percent, placing it in the top five after Banten, West Java, Maluku, East Kalimantan, and West Papua as the regions with the highest unemployment. As a result, a study was conducted to investigate the factors suspected of being the cause of the open unemployment rate (TPT) in South Sulawesi in 2020. The purpose of this study is to look into the factors that are thought to affect the TPT in South Sulawesi. Based on the scatterplot, it is determined that there is no discernible pattern between Open Unemployment Rate (TPT) and the factors thought to have an effect, so the spline truncated nonparametric regression method is applied. According to the study results, the best nonparametric spline truncated regression model for modeling Open Unemployment Rate (TPT) in South Sulawesi has three knot points. The resulting model has a minimum GCV of 2.16 and an R^2 of 98.21%. According to the results of simultaneous and partial tests, the percentage of the poor, average length of schooling, dependency ratio, and human development index all had a significant effect on the open unemployment rate in South Sulawesi in 2020.

Keywords: Nonparametric Regression, Spline Truncated, Open Unemployment Rate (TPT)

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN (TUGAS AKHIR)	ii
PERNYATAAN KEASLIAN	Error! Bookmark not defined.
KATA PENGANTAR	iv
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Definisi Pengangguran	6
2.2 Statistika deskriptif.....	9
2.3 Analisis Regresi	9
2.4 <i>Likelihood Ratio Test (LRT)</i>	11
2.5 Model Regresi Nonparametrik dengan <i>Spline Truncated</i>	12
2.7 Rumusan Pengujian Hipotesis pada Regresi Nonparametrik <i>Spline Truncated</i>	17
2.8 Pengujian Hipotesis Parameter Model Regresi Nonparametrik <i>Spline Truncated</i>	27
2.9 Kriteria Keباikan Model.....	29
2.10 Pemeriksaan Asumsi Residual	30

2.11	Teorema Dasar Terkait Dengan Aljabar Matriks	32
BAB III METODE PENELITIAN.....		34
3.1	Sumber Data	34
3.2	Deskripsi Variabel.....	34
3.3	Langkah Penelitian.....	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Eksplorasi Data.....	38
4.2	Pemodelan TPT Provinsi Sulawesi Selatan Menggunakan Regresi Nonparametrik Spline <i>Truncated</i>	41
4.3	Pemilihan Titik Knot Optimal	41
4.4	Penaksiran Parameter Tingkat Pengangguran Terbuka di Provinsi Sulawesi Selatan Tahun 2020	49
4.5	Pengujian Parameter Model Regresi	50
4.6	Pengujian Asumsi Residual	52
BAB V.....		60
PENUTUP.....		60
5.1	Kesimpulan	60
5.2	Saran	61
LAMPIRAN		65

DAFTAR TABEL

Tabel 2.1 Analisis of Varians (ANOVA).....	28
Tabel 3.1 Struktur Data.....	36
Tabel 4. 1 Statistik Deskriptif Keseluruhan Data.....	38
Tabel 4. 2 Nilai GCV dengan Satu Titik Knot Spline <i>Truncated</i> Linear.....	42
Tabel 4. 3 Nilai GCV dengan Dua Titik Knot Spline <i>Truncated</i> Linear	44
Tabel 4. 4 Nilai GCV dengan Tiga Titik Knot Spline <i>Truncated</i> Linear	46
Tabel 4.5 Perbandingan Nilai GCV Minimum	48
Tabel 4.6 Estimasi Model Spline Linier Tiga Titik Knot	49
Tabel 4.7 ANOVA Model Regresi Spline <i>Truncated</i>	50
Tabel 4.8 Uji Parsial Model Regresi Nonparametrik Spline <i>Truncated</i>	51
Tabel 4. 9 ANOVA dari Uji Glejser	52

DAFTAR GAMBAR

Gambar 4.1 *Scatter plot* antara variabel respon dan variabel prediktor pada data tingkat pengangguran terbuka tahun 2020..... 40

Gambar 4.2 Estimasi Kurva Spline Linear Satu Titik Knot..... 43

Gambar 4.3 Estimasi Kurva Spline Linear Dua Titik Knot 45

Gambar 4.4 Estimasi Kurva Spline Linear Tiga Titik Knot 48

Gambar 4.5 ACF Residual 53

Gambar 4.6 Uji Normalitas 54

DAFTAR LAMPIRAN

Lampiran 1: Data Penelitian.....66
Lampiran 2: Data Variabel Prediktor Tingkat Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 202068

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan salah satu negara berkembang yang dalam pengelompokan negara berdasarkan taraf kesejahteraan masyarakatnya, salah satu permasalahan yang dihadapi sebagai negara berkembang adalah masalah pengangguran (Muslim, 2014). Pengangguran merupakan masalah yang sangat kompleks karena mempengaruhi sekaligus dipengaruhi oleh banyak faktor yang saling berinteraksi mengikuti pola yang tidak selalu mudah untuk dipahami.

Indikator utama yang digunakan untuk mengukur angka pengangguran dalam angkatan kerja yaitu Tingkat Pengangguran Terbuka (BPS, 2009). Pengangguran terbuka merupakan salah satu jenis pengangguran dimana angkatan kerja tidak memiliki pekerjaan sama sekali. Badan Pusat Statistik menjelaskan bahwa pengangguran terbuka terdiri dari mereka yang tak punya pekerjaan dan mencari pekerjaan, mereka yang tidak punya pekerjaan dan sedang mempersiapkan usaha, mereka yang tak punya pekerjaan dan tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan, serta mereka yang sudah punya pekerjaan tetapi belum mulai bekerja. Pengangguran terbuka dapat diukur dengan presentase pengangguran terhadap angkatan kerja, hasil perhitungan tersebut dinamakan tingkat pengangguran terbuka (TPT).

Sulawesi Selatan menjadi salah satu daerah penyumbang penganggur tertinggi di Indonesia. Pada Februari 2020 tercatat angka pengangguran di Sulawesi Selatan sebesar 6,07 persen menjadikan Sulawesi Selatan terdata masuk lima besar setelah Banten, Jawa Barat, Maluku, Kalimantan Timur dan Papua Barat sebagai daerah dengan pengangguran tertinggi. Jumlah pengangguran di Sulawesi Selatan pada Februari 2020 dibandingkan Agustus 2020 terjadi peningkatan yakni mencapai 6,31 persen.

Salah satu upaya yang dapat dilakukan untuk mengurangi pengangguran adalah melakukan penanganan terhadap faktor-faktor yang mempengaruhi TPT. Banyak metode yang bisa digunakan dalam menganalisis faktor-faktor yang menyebabkan TPT. Salah satunya adalah analisis regresi. Analisis regresi bertujuan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor. Dalam analisis regresi terdapat tiga pendekatan dalam menganalisis kurva regresi yaitu pendekatan parametrik, nonparametrik, dan semiparametrik. Regresi parametrik digunakan jika hubungan variabel respon dan variabel prediktor memiliki pola tertentu atau membentuk pola data yang jelas seperti linier, kuadratik, atau kubik. Pada kenyataannya antara variabel respon dan variabel prediktor tidak selalu memiliki pola hubungan yang jelas. Metode yang dapat digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor yang kurva regresinya tidak diketahui polanya ataupun jika bentuk polanya berubah pada tiap sub interval tertentu adalah regresi nonparametrik. Model regresi nonparametrik yang sering mendapat perhatian dari para peneliti adalah *Kernel* (Hardle, 1990), *Spline* (Wahba, 1990), Deret *Fourier* (Antoniadis A. G., 1994) dan *Wavelet* (Antoniadis A. , 2007).

Pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi karena data diharapkan bisa mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektifitas peneliti. Salah satu pendekatan yang dapat digunakan dalam mengestimasi model regresi nonparametrik adalah spline. Model spline memiliki interpretasi statistik dan interpretasi visual yang sangat baik serta fleksibilitas yang tinggi (Eubank, 1999). Beberapa jenis fungsi spline yang telah diteliti sebelumnya antara lain, *smoothing spline* (Eubank, 1999), B-Spline (Lyche, 2008), *penalized spline* (Griggs, 2013), *thin plate spline* (Wood, 2003), dan sebagainya. Budiantara (2005) mengembangkan estimator spline dalam regresi nonparametrik dengan menggunakan basis fungsi keluarga spline *truncated*. Fungsi spline *truncated* merupakan fungsi polinomial yang terpotong-potong pada suatu titik knot. Titik knot merupakan titik perpaduan bersama dimana fungsi tersebut

terpotong, atau titik yang menggambarkan terjadinya perubahan perilaku data pada sub-sub interval tertentu (Budiantara I. , 2009). Oleh karena itu, model spline *truncated* memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Budiantara I., 2011). Sementara itu, Wahba (1990) memberikan metode untuk memilih parameter penghalus optimal dalam estimator spline yaitu dengan *Generalized Crossn Validation* (GCV).

Penelitian sebelumnya mengenai TPT telah dilakukan oleh Arjun (2019) yang melakukan pemodelan faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di Kalimantan menggunakan regresi nonparametrik spline *truncated*. Hasil penelitian menunjukkan bahwa faktor-faktor yang berpengaruh signifikan terhadap tingkat pengangguran terbuka di Kalimantan adalah rata-rata lama sekolah, tingkat partisipasi angkatan kerja, persentase penduduk miskin, laju pertumbuhan ekonomi, dan jumlah industri besar sedang. Wijaya (2018) juga meneliti tentang faktor-faktor yang berpengaruh terhadap TPT di Provinsi Aceh menggunakan regresi nonparametrik spline *truncated*. Hasil penelitian menunjukkan menunjukkan bahwa variabel *dependency ratio*, dan presentase penduduk miskin berpengaruh secara signifikan terhadap TPT di Provinsi Aceh tahun 2015. Amalia & Sari (2019) melakukan penelitian terhadap TPT di pulau Jawa tahun 2017 menggunakan analisis spasial. Hasil penelitian menunjukkan bahwa indeks pembangunan manusia, *dependency ratio*, tingkat partisipasi angkatan kerja, dan upah minimum kabupaten/kota berpengaruh signifikan terhadap TPT di Pulau Jawa tahun 2017.

Berdasarkan paparan dari penelitian-penelitian yang telah dilakukan sebelumnya menggunakan pendekatan regresi. Peneliti tertarik untuk memodelkan kasus tingkat pengangguran terbuka di Provinsi Sulawesi Selatan sebagai provinsi dengan kontribusi pengangguran yang cukup tinggi di Indonesia tahun 2020 menggunakan pendekatan regresi nonparametrik. Dengan alasan, studi awal yang dilakukan peneliti terlihat bahwa tidak adanya pola tertentu antara variabel respon

dengan variabel prediktor-prediktornya sehingga akan lebih tepat jika didekati menggunakan pendekatan regresi nonparametrik dengan estimator spline *truncated* sebagai regresi nonparametrik yang cukup populer digunakan. Penelitian ini menggunakan variabel berdasarkan penelitian-penelitian yang pernah dilakukan sebelumnya. Selanjutnya akan dilakukan pengujian hipotesis terhadap model regresi nonparametrik guna mengetahui apakah variabel prediktor yang digunakan memberikan pengaruh atau tidak terhadap tingkat pengangguran terbuka.

Pengujian hipotesis merupakan suatu metode untuk mengestimasi parameter populasi dengan cara menguji kebenaran dari suatu pernyataan. Melalui pengujian hipotesis akan diperoleh bentuk dari hipotesis serta statistik uji dan distribusinya. Sejumlah penelitian sebelumnya tentang pengujian hipotesis antara lain Husni (2018) melakukan pengujian hipotesis parsial untuk parameter model regresi nonparametrik spline *truncated* multivariabel terhadap data kematian ibu di Provinsi Nusa Tenggara Timur. Ferdiana (2017) juga melakukan kajian tentang pengujian hipotesis simultan dalam regresi semiparametrik spline *truncated*.

Berdasarkan uraian tersebut, penulis akan membahas tentang pengujian hipotesis model regresi nonparametrik spline *truncated* dalam pemodelan tingkat pengangguran terbuka di Sulawesi Selatan tahun 2020.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini yaitu bagaimana pemodelan regresi nonparametrik spline *truncated* terhadap Tingkat Pengangguran Terbuka (TPT) di Sulawesi Selatan.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan sebelumnya, tujuan penelitian ini yaitu untuk memperoleh model regresi nonparametrik spline *truncated* terhadap Tingkat Pengangguran Terbuka (TPT) di Sulawesi Selatan tahun 2020.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah dapat memberikan informasi mengenai faktor-faktor yang berpengaruh terhadap tingkat pengangguran terbuka serta dapat digunakan sebagai referensi pada penelitian-penelitian selanjutnya yang berhubungan dengan estimator spline *truncated*.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan adalah data indikator tingkat pengangguran terbuka Provinsi Sulawesi Selatan tahun 2020.
2. Pemilihan titik knot optimal menggunakan metode GCV (*Generalized Cross Validation*).
3. Jumlah titik knot yang digunakan dalam pemodelan sebanyak 1, 2, dan 3 titik knot.

BAB II TINJAUAN PUSTAKA

2.1 Definisi Pengangguran

Pengangguran merupakan hal yang akan selalu muncul didalam perekonomian, dimana saat pengeluaran agregatnya lebih rendah dibandingkan dengan kemampuan faktor-faktor produksi yang telah tersedia didalam perekonomian untuk dapat menghasilkan barang-barang dan juga jasa (Prasaja, 2011). Navarrete menjelaskan dalam bukunya “*Underemployment in Underdeveloped Countries*” pengangguran dapat dilukiskan sebagai suatu keadaan dimana adanya pengalihan sejumlah faktor tenaga kerja ke bidang lain yang mana tidak akan mengurangi output keseluruhan sektor asalnya atau dikatakan bahwa peoduktivitas marginal unit-unit faktor tenaga tempat asal mereka bekerja adalah nol atau hampir mendekati nol atau juga negatif (Jhingan, 2014).

Salah satu alasan pengangguran selalu muncul didalam pengangguran adalah pencarian kerja. Pencarian kerja (*job search*) adalah suatu proses seseorang untuk mencocokkan pekerja dengan pekerjaan yang sesuai dengan bakat dan juga keterampilan sesuai yang dimiliki oleh mereka. Namun, jika semua pekerja dan pekerjaan tidak ada bedanya, maka tidak menutup kemungkinan bagi para pekerja bahwa mereka cocok dengan pekerjaan apa saja, akan tetapi pada kenyataannya bakat dan juga kemampuan seseorang itu berbeda-beda (Mankiw, 2012).

Definisi pengangguran menurut BPS pengangguran terbuka (*open unemployment*) didasarkan pada konsep seluruh angkatan kerja yang mencari perkerjaan, baik yang mencari perkerjaan pertama kali maupun yang pernah bekerja sebelumnya.

a. Pengangguran dalam Sektor Informal

Pengangguran terbuka biasanya terjadi pada generasi muda yang baru menyelesaikan pendidikan menengah dan tinggi. Ada kecenderungan mereka yang baru menyelesaikan pendidikan berusaha mencari kerja sesuai dengan aspirasi

mereka. Aspirasi mereka biasanya adalah bekerja di sektor modern atau di kantor, untuk mendapatkan pekerjaan itu mereka bersedia menunggu untuk beberapa lama, tidak tertutup kemungkinan mereka berusaha mencari pekerjaan itu di kota atau di provinsi atau daerah yang kegiatan industri telah berkembang. Hal ini menyebabkan angka pengangguran tinggi di perkotaan atau di daerah kegiatan industri atau sektor modern berkembang. Sebaliknya pengangguran terbuka rendah di daerah atau provinsi yang tumpu pada sektor pertanian. Hal tersebut penyediaan pekerjaan di sektor informal oleh sebab rendahnya pendidikan dan kurang menjamin kelangsungan hidup.

b. Pengukuran Tingkat Pengangguran

Badan statistik negara mengelompokkan orang dewasa pada setiap rumah tangga yang disurvei ke dalam satu kategori berikut.

1. Bekerja
2. Pengangguran
3. Tidak termasuk angkatan kerja

Setelah mengelompokkan seluruh individu yang disurvei ke dalam tiga kategori tersebut, badan statistik negara menghitung berbagai statistik untuk merangkum kondisi angkatan kerja. Angkatan kerja (*labor force*) adalah jumlah orang yang berkerja dan tidak berkerja.

$$\text{Angkatan kerja} = \text{Jumlah orang yang bekerja} + \text{Jumlah yang tidak bekerja}$$

Tingkat pengangguran (*unemployment rate*) adalah persentase angkatan kerja yang tidak bekerja:

$$\text{Tingkat Pengangguran} = \frac{\text{jumlah pengangguran}}{\text{jumlah angkatan kerja}} \times 100\%$$

Setelah itu, tingkat pengangguran untuk seluruh populasi penduduk dewasa dan untuk kelompok yang lebih sempit seperti laki-laki dan perempuan dapat dihitung.

2.1.1 Variabel Prediktor yang Mempengaruhi Tingkat Pengangguran Terbuka

Variabel-variabel yang berpengaruh terhadap TPT menurut penelitian sebelumnya antara lain Persentase Penduduk Miskin, Rata-rata Lama Sekolah, *Dependency Ratio*, dan Indeks Pembangunan Manusia.

Persentase penduduk miskin merupakan salah satu faktor yang mempengaruhi TPT. Presentase penduduk miskin adalah presentase penduduk yang berada dibawah garis kemiskinan. Garis kemiskinan merupakan jumlah nilai pengeluaran dari 52 komoditi dasar makanan yang riil dikonsumsi penduduk referensi yang kemudian disetarakan dengan 2100 kilokalori perkapita perhari dengan nilai kebutuhan minimum dari komoditi-komoditi non-makanan terpilih yang meliputi perumahan, sandang, pendidikan dan kesehatan (BPS, 2017).

Dependency Ratio juga mempengaruhi tingkat pengangguran terbuka. Rasio Ketergantungan (*Dependency Ratio*) adalah perbandingan antara jumlah penduduk umur 0-14 tahun, ditambah dengan jumlah penduduk 65 tahun ke atas (keduanya disebut dengan bukan angkatan kerja) dibandingkan dengan jumlah penduduk usia 15-64 tahun (angkatan kerja). Rasio ketergantungan (*dependency ratio*) dapat digunakan sebagai indikator yang secara kasar dapat menunjukkan keadaan ekonomi suatu Negara apakah tergolong negara maju atau negara yang sedang berkembang, (BPS, 2016). Jika jumlah pengangguran tinggi maka rasio ketergantungan tinggi pula dikarenakan negara memiliki tanggungan yang besar untuk penduduk dimana kondisi tersebut mampu menghambat pembangunan dan masalah sosial lainnya

Faktor lain yang juga mempengaruhi tingkat pengangguran terbuka adalah indeks pembangunan manusia. Burhanuddin (2015) dalam penelitiannya mengenai hubungan indeks pembangunan manusia dengan tingkat pengangguran bahwa menyimpulkan bahwa indeks pembangunan manusia memiliki pengaruh yang signifikan dan negatif terhadap tingkat pengangguran. Hal ini menjelaskan bahwa semakin tinggi angka indeks pembangunan manusia pada suatu wilayah maka akan menyebabkan tingkat pengangguran semakin menurun dan sebaliknya indeks

pembangunan manusia rendah akan berdampak pada tingginya tingkat pengangguran di wilayah tersebut.

2.2 Statistika deskriptif

Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan, penyusunan dan penyajian suatu gugus data serta penarikan kesimpulan sehingga memberikan informasi yang berguna (Walpole, 1995). Statistika deskriptif bertugas untuk menggambarkan (*description*) tentang suatu gejala. Statistika deskriptif hanya memberikan informasi mengenai data itu sendiri dan sama sekali tidak menarik inferensia atau kesimpulan apapun dari gugus data induknya yang lebih besar. Informasi yang dapat diperoleh dari statistika deskriptif ini antara lain ukuran pemusatan data, ukuran penyebaran data, serta kecenderungan suatu gugus data. Pengukuran pemusatan data dilakukan dengan menghitung nilai rata-rata (*mean*) dan pengukuran penyebaran data dilakukan dengan menghitung nilai standar deviasi. Rumus untuk perhitungan rata-rata (*mean*) dari x_1, x_2, \dots, x_n data yaitu sebagai berikut:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.1)$$

dengan

\bar{x} : rata-rata (*mean*)

x_i : pengamatan ke- i , $i = 1, 2, \dots, n$

n : banyaknya pengamatan n

2.3 Analisis Regresi

Analisis regresi digunakan untuk mengetahui pengaruh dari suatu variabel terhadap variabel lain. Analisis regresi merupakan sebuah metode statistika yang memberikan penjelasan tentang pola hubungan (model) antara dua variabel atau lebih (Drapper & dan Smith, 1992). Dalam analisis regresi, variabel yang mempengaruhi adalah variabel bebas (variabel prediktor) dan variabel yang dipengaruhi adalah

variabel terikat (variabel respon). *Scatter plot* sering kali digunakan dalam mempelajari pola hubungan antara variabel respon dengan variabel prediktor. Plot dapat menunjukkan apakah kurva membentuk suatu pola linier, kuadratik, ataupun kubik. Akan tetapi pada kenyataannya kurva yang dihasilkan sering kali tidak bisa ditentukan hanya dengan melihat bentuk polanya secara visual. Oleh sebab itu dalam analisis regresi terdapat dua pendekatan yang sering digunakan untuk mengestimasi kurva yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik.

2.3.1 Regresi Parametrik

Regresi parametrik merupakan metode yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor dengan asumsi bahwa bentuk kurva regresinya diketahui. Secara matematis, bentuk regresi parametrik dapat ditulis dalam persamaan (2.2).

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.2)$$

Fungsi $g(x_i)$ dapat ditulis dalam persamaan (2.3) sebagai berikut.

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i \quad (2.3)$$

dengan $\mathbf{x}'_i = [1 \quad x_{1i} \quad x_{2i} \quad \dots \quad x_{pi}]$, $i = 1, 2, \dots, n$; sedangkan n adalah banyaknya data dan p adalah banyak variabel, sementara $\boldsymbol{\beta}' = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_p]$. Sehingga persamaannya menjadi persamaan (2.4).

$$y_i = [1 \quad x_{1i} \quad x_{2i} \quad \dots \quad x_{pi}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon_i \quad (2.4)$$

$$= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

dengan

y_i : variabel respon

$g(x_i)$: fungsi kurva regresi parametrik

$x_{1i}, x_{2i}, \dots, x_{pi}$: variabel prediktor ke- k pada pengamatan ke- i

β_0 : intersep

- β_p : parameter regresi ke- k
 ε_i : sisaan pada pengamatan ke- i yang diasumsikan identik, independen, dan berdistribusi $N(0, \sigma^2)$.

2.3.2 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. Kurva regresi pada regresi nonparametrik diasumsikan fungsi *smooth* (mulus) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga regresi nonparametrik memiliki sifat fleksibilitas yang tinggi (Eubank R. , 1988). Bentuk umum model regresi nonparametrik dapat dituliskan sebagai berikut:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.5)$$

dimana

- y_i : variabel respon
 f : fungsi regresi nonparametrik
 x_{pi} : variabel prediktor ke- p ($p = 1, 2, \dots, j$) pada pengamatan ke- i
 ε_i : sisaan pada pengamatan ke- i yang diasumsikan identik, independen, dan berdistribusi $N(0, \sigma^2)$.

2.4 Likelihood Ratio Test (LRT)

Metode *likelihood ratio* pada pengujian hipotesis berkaitan dengan *Maximum Likelihood Estimation* (MLE). Misalkan x_1, x_2, \dots, x_n adalah sampel random dari populasi dengan *probability density function* (pdf), maka fungsi *likelihood* didefinisikan sebagai berikut (Casella dan Berger, 2001):

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$$

$$\begin{aligned}
&= \prod_{i=1}^n f(x_i, \theta) \\
&= L(\theta | x_1, x_2, \dots, x_n) \\
&= L(\theta | x) \tag{2.6}
\end{aligned}$$

Kemudian Persamaan (2.6) didiferensialkan terhadap θ untuk memperoleh penaksiran yang maksimum. Dalam banyak kasus, penggunaan diferensiasi akan lebih mudah bekerja pada logaritma natural dari $L(x_1, x_2, \dots, x_n; \theta)$, yaitu:

$$\ln L(x_1, x_2, \dots, x_n; \theta) \tag{2.7}$$

Langkah-langkah untuk menentukan penaksiran maksimum likelihood dari θ_i adalah:

1. Menentukan fungsi likelihood

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$$

2. Membentuk logaritma natural likelihood

$$\ln L(x_1, x_2, \dots, x_n; \theta) = \ln f(x_1; \theta) + \ln f(x_2; \theta) + \dots + \ln f(x_n; \theta)$$

3. Menurunkan persamaan logaritma natural likelihood terhadap θ

$$\frac{\partial \ln L(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} = 0$$

4. Didapat penaksiran maksimum likelihood θ

2.5 Model Regresi Nonparametrik dengan *Spline Truncated*

Diantara beberapa model regresi nonparametrik, spline merupakan model yang mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik (Budiantara, 2009). Salah satu bentuk fungsi spline adalah spline *truncated*. Dalam spline *truncated* terdapat dua komponen yaitu komponen polinomial dan komponen *truncated*. Salah satu kelebihan spline *truncated* adalah model ini mengikuti pola data sesuai pergerakannya dengan adanya titik titik knot.

Titik knot adalah titik yang menunjukkan perubahan data pada sub-sub interval (Budiantara, 2009). Secara umum, bentuk fungsi spline *truncated* derajat q dengan titik titik knot K_1, K_2, \dots, K_r dapat dinyatakan dalam persamaan (2.8),

$$f(x_i) = \sum_{l=0}^q \beta_l x_i^l + \sum_{h=1}^r \beta_{q+h} (x_i - K_h)_+^q + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.8)$$

dengan $\beta_0, \beta_1, \dots, \beta_q, \beta_{q+1}, \dots, \beta_{q+r}$ adalah parameter regresi, K_h adalah titik knot ke- h , ($h = 1, 2, \dots, r$) dan $(x_i - K_h)_+^q$ adalah fungsi polinomial *truncated* dengan:

$$(x_i - K_h)_+^q = \begin{cases} (x_i - K_h)^q, & \text{jika } x_i \geq K_h \\ 0, & \text{jika } x_i < K_h \end{cases}$$

Jika dalam persamaan (2.8) disubstitusikan nilai $q = 1, 2, 3$ maka diperoleh fungsi spline yang berturut-turut dinamakan spline linear, spline kuadrat, dan spline kubik (Rodriguez, 2001).

Fungsi yang menyatakan hubungan antara prediktor ke- p dengan respon tunggal jika dihipotesiskan dengan fungsi spline $f(x_i)$ dalam persamaan (2.8) maka dapat dinyatakan dalam bentuk sebagai berikut:

$$\begin{aligned} y_i &= f(x_{1i}) + f(x_{2i}) + \dots + f(x_{pi}) + \varepsilon_i \\ &= \sum_{j=1}^p f(x_{ji}) + \varepsilon_i; \quad i = 1, 2, \dots, n \end{aligned} \quad (2.9)$$

dengan:

$$f(x_{ji}) = \beta_{0j} + \sum_{l=1}^q \beta_{jl} x_{ji}^l + \sum_{h=1}^r \beta_{j(q+h)} (x_{ji} - K_{jh})_+^q; \quad j = 1, 2, \dots, p$$

Sisaan (ε_i) pada pengamatan ke- i , yang diasumsikan saling bebas berdistribusi normal dengan nol dan varians σ^2 , y_i adalah variabel respon pada pengamatan ke- i , x_{ji} adalah variabel prediktor ke- j pada pengamatan ke- i , β_{0j} adalah intersep prediktor ke- j , β_{jl} adalah parameter polinomial pada prediktor ke- j dan orde ke- l , $\beta_{j(q+h)}$ adalah parameter *truncated* pada prediktor ke- j dan titik knot ke- $(q + h)$, K_{jh} adalah nilai titik knot pada prediktor ke- j dan titik knot ke- h , r adalah banyaknya titik knot, q adalah orde polinomial spline *truncated*, dan p adalah banyaknya variabel prediktor.

Persamaan (2.9) untuk n data pengamatan dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.10)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21}^2 & \cdots & x_{p1}^q & (x_{11} - K_{11})_+^q & \cdots & (x_{p1} - K_{pr})_+^q \\ 1 & x_{12} & x_{22}^2 & \cdots & x_{p2}^q & (x_{12} - K_{11})_+^q & \cdots & (x_{p2} - K_{pr})_+^q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n}^2 & \cdots & x_{pn}^q & (x_{1n} - K_{11})_+^q & \cdots & (x_{pn} - K_{pr})_+^q \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \\ \beta_{(q+1)} \\ \vdots \\ \beta_{(q+r)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Pada model (2.10), \mathbf{y} menyatakan vektor kolom untuk variabel respons berukuran $n \times 1$, \mathbf{X} adalah matriks konstanta berukuran $n \times (1 + q + r)$, $\boldsymbol{\beta}$ adalah vector parameter berukuran $(1 + q + r) \times 1$. Apabila diasumsikan $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$, karena \mathbf{y} merupakan kombinasi linier dari $\boldsymbol{\varepsilon}$ maka \mathbf{y} juga berdistribusi normal $\mathbf{y} \sim N(E(\mathbf{y}), Var(\mathbf{y}))$.

Persamaan (2.10) dapat dituliskan dalam bentuk sebagai berikut:

$$\mathbf{y} = \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta} + \boldsymbol{\varepsilon} ,$$

Selanjutnya nilai dari $E(\mathbf{y})$ dan $Var(\mathbf{y})$ diuraikan dengan cara sebagai berikut:

$$\begin{aligned} E(\mathbf{y}) &= E(\mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \mathbf{X}[K]\boldsymbol{\beta} + E(\boldsymbol{\varepsilon}) \\ &= \mathbf{X}[K]\boldsymbol{\beta} \end{aligned} \quad (2.11)$$

$$\begin{aligned} Var(\mathbf{y}) &= Var(\mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \mathbf{0} + Var(\boldsymbol{\varepsilon}) \\ &= \sigma^2 \mathbf{I} \end{aligned} \quad (2.12)$$

Berdasarkan (2.11) dan (2.12) didapatkan \mathbf{y} berdistribusi normal dengan *mean* $\mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta}$ dan *varians* $\sigma^2 \mathbf{I}$. Selanjutnya, estimasi titik dari $\boldsymbol{\beta}$ akan didapatkan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Distribusi probabilitas dari $\boldsymbol{\varepsilon}$ adalah:

$$\begin{aligned}
 f(\boldsymbol{\varepsilon}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}\right) \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta})\right) \quad (2.13)
 \end{aligned}$$

Berdasarkan persamaan (2.13) diperoleh fungsi *likelihood* seperti persamaan dibawah ini:

$$\begin{aligned}
 L(\boldsymbol{\beta}) &= \prod_{i=1}^n f(\boldsymbol{\varepsilon}_i) \\
 &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta})\right) \quad (2.14)
 \end{aligned}$$

Apabila persamaan (2.14) dibuat transformasi logaritma, akan didapatkan

$$\begin{aligned}
 l(\boldsymbol{\beta}) &= \log L(\boldsymbol{\beta}) \\
 &= \log\left((2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}[K_1, K_2, \dots, K_r]\boldsymbol{\beta})\right)\right) \\
 &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}) \quad (2.15)
 \end{aligned}$$

Dengan menggunakan derivatif parsial terhadap $\boldsymbol{\beta}$ diperoleh:

$$\begin{aligned}
 \frac{\partial l(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial\left(-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} \\
 &= \frac{\partial\left(-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}[K]'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} \\
 &= -\frac{1}{2\sigma^2} (-2\mathbf{X}[K]'\mathbf{y} + 2\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta}) \quad (2.16)
 \end{aligned}$$

Jika derivatif parsial tersebut disamakan dengan nol, akan diperoleh persamaan:

$$2\mathbf{X}[K]'\mathbf{X}[K]\hat{\boldsymbol{\beta}} = 2\mathbf{X}[K]'\mathbf{y}$$

Sehingga estimator parameter $\boldsymbol{\beta}$ adalah:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}[K]'\mathbf{X}[K])^{-1} \mathbf{X}[K]'\mathbf{y} \quad (2.17)$$

Selanjutnya menentukan estimator model regresi nonparametrik spline *truncated*.

$$\begin{aligned}
 \hat{\mathbf{y}} &= \mathbf{X}[K]\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
 &= \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1} \mathbf{X}[K]'\mathbf{y} + \boldsymbol{\varepsilon}
 \end{aligned}$$

$$= \mathbf{A}\mathbf{y} + \boldsymbol{\varepsilon} \quad (2.18)$$

dengan $\mathbf{A} = \mathbf{X}[\mathbf{K}](\mathbf{X}[\mathbf{K}]'\mathbf{X}[\mathbf{K}])^{-1}\mathbf{X}[\mathbf{K}]'$ adalah matriks yang digunakan untuk perhitungan pada rumus *Generalized Cross Validation* (GCV) dalam pemilihan titik knot.

2.6 Pemilihan Titik Knot Optimal

Dalam regresi nonparametrik spline *truncated*, hal penting yang berperan dalam mendapatkan estimator spline *truncated* terbaik adalah pemilihan titik knot yang optimal. Jika dalam sekumpulan data jumlah titik knot yang dipilih terlalu sedikit, maka akan menghasilkan estimasi kurva regresi yang sangat global dan menyebabkan biasanya lebih besar. Sebaliknya, jika dalam sekumpulan data jumlah knot yang dipilih terlalu banyak, maka akan menyebabkan estimasi kurva regresi sangat kasar. Oleh sebab itu dibutuhkan suatu metode yang dapat memilih titik knot optimal.

Salah satu metode yang sering digunakan dalam memilih titik knot optimal adalah *Generalized Cross Validation* (GCV). Menurut Wahba (1990) jika dibandingkan dengan metode lain, misalnya *Cross Validation* (CV) dan metode *Unbiased Risk* (UBR) ataupun *Generalized Maximum Likelihood* (GML), GCV secara teoritis memiliki sifat optimal asimtotik atau dengan sampel besar yang sifatnya tetap optimal. Wahba (1990) juga menyatakan bahwa metode GCV juga memiliki kelebihan tidak memerlukan pengetahuan terhadap variansi populasi serta metode GCV invariants terhadap transformasi.

Fungsi GCV untuk pemilihan titik knot optimal pada regresi nonparametrik ditunjukkan dalam persamaan (2.19).

$$GCV(K) = \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{[n^{-1} \text{trace}(\mathbf{I} - \mathbf{A}(K))]^2} \quad (2.19)$$

dimana y_i adalah variabel respon, \hat{y}_i adalah nilai estimasi variabel respon, $i = 1, 2, \dots, n$ yang merupakan jumlah observasi, K_h merupakan titik titik knot, \mathbf{I} adalah

matriks identitas, dan matriks $\mathbf{A} = \mathbf{X}[\mathbf{K}](\mathbf{X}[\mathbf{K}]'\mathbf{X}[\mathbf{K}])^{-1}\mathbf{X}[\mathbf{K}]'$. Nilai titik knot yang optimal diberikan oleh nilai GCV minimum (Budiantara, 2006).

2.7 Rumusan Pengujian Hipotesis pada Regresi Nonparametrik Spline Truncated

Diberikan model regresi nonparametrik *spline truncated* derajat q dengan knot-knot K_1, K_2, \dots, K_r .

$$y_i = \beta_{0j} + \sum_{l=1}^q \beta_{jl} x_{ji}^l + \sum_{h=1}^r \beta_{j(q+h)} (x_{ji} - K_{jh})_+^q + \varepsilon_i; i = 1, 2, \dots, n, j = 1, 2, \dots, p$$

Untuk menurunkan uji hipotesis H_0 lawan H_1 dapat menggunakan metode LRT. Perhatikan model regresi spline, dengan ε_i berdistribusi independen identik $N(0, \sigma^2 \mathbf{I})$. Karena error $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$, maka $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ fungsi *likelihood* diberikan oleh:

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= \prod_{i=1}^n \left((2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - f(x_{ji}))^2\right) \right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_{ji}))^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}[\mathbf{K}]\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}[\mathbf{K}]\boldsymbol{\beta})\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}[\mathbf{K}]'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}[\mathbf{K}]'\mathbf{X}[\mathbf{K}]\boldsymbol{\beta})\right) \end{aligned}$$

dengan memperhatikan parameter yang terdapat di bawah $H(\Omega)$, diperoleh fungsi *likelihood* sebagai berikut:

$$\begin{aligned} L(\boldsymbol{\beta}_\Omega, \sigma_\Omega^2) &= (2\pi\sigma_\Omega^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_\Omega^2} (\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'_\Omega \mathbf{X}[\mathbf{K}]'\mathbf{y} + \boldsymbol{\beta}'_\Omega \mathbf{X}[\mathbf{K}]'\mathbf{X}[\mathbf{K}]\boldsymbol{\beta}_\Omega)\right) \\ &= -\frac{n}{2} \log(2\pi\sigma_\Omega^2) - \frac{1}{2\sigma_\Omega^2} (\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'_\Omega \mathbf{X}[\mathbf{K}]'\mathbf{y} + \boldsymbol{\beta}'_\Omega \mathbf{X}[\mathbf{K}]'\mathbf{X}[\mathbf{K}]\boldsymbol{\beta}_\Omega) \quad (2.20) \end{aligned}$$

Selanjutnya mencari turunan dari persamaan (2.20) terhadap $\boldsymbol{\beta}_\Omega$ untuk mendapatkan estimasi parameter $\boldsymbol{\beta}_\Omega$ sebagai berikut:

$$\frac{\partial \log L(\boldsymbol{\beta}_\Omega, \sigma_\Omega^2)}{\partial \boldsymbol{\beta}_\Omega} = \frac{\partial}{\partial \boldsymbol{\beta}_\Omega} \left[-\frac{n}{2} \log(2\pi\sigma_\Omega^2) - \frac{1}{2\sigma_\Omega^2} (\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'_\Omega \mathbf{X}[\mathbf{K}]'\mathbf{y} + \boldsymbol{\beta}'_\Omega \mathbf{X}[\mathbf{K}]'\mathbf{X}[\mathbf{K}]\boldsymbol{\beta}_\Omega) \right]$$

$$\begin{aligned}
 &= 0 - \frac{1}{2\sigma_\Omega^2}(-2\mathbf{X}[K]'\mathbf{y} + 2\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta}_\Omega) \\
 &\quad - \frac{1}{2\sigma_\Omega^2}(-2\mathbf{X}[K]'\mathbf{y} + 2\mathbf{X}[K]'\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega) = 0
 \end{aligned}$$

$$2\mathbf{X}[K]'\mathbf{y} = 2\mathbf{X}[K]'\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega$$

$$\widehat{\boldsymbol{\beta}}_\Omega = (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \quad (2.21)$$

dan untuk mendapatkan estimasi parameter σ_Ω^2 dengan mencari turunan dari persamaan (2.20) terhadap σ_Ω^2 sebagai berikut:

$$\begin{aligned}
 \frac{\partial \log L(\boldsymbol{\beta}_\Omega, \sigma_\Omega^2)}{\partial \sigma_\Omega^2} &= \frac{\partial}{\partial \sigma_\Omega^2} \left[-\frac{n}{2} \log(2\pi\sigma_\Omega^2) - \frac{1}{2\sigma_\Omega^2} ((\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}_\Omega)) \right] \\
 &= -\frac{n}{2} \left(\frac{2\pi}{2\pi\sigma_\Omega^2} \right) + \frac{1}{2(\sigma_\Omega^2)^2} (\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}_\Omega) \\
 \frac{n}{2\hat{\sigma}_\Omega^2} &= \frac{1}{2(\hat{\sigma}_\Omega^2)^2} (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega) \\
 \hat{\sigma}_\Omega^2 &= \frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)}{n} \quad (2.22)
 \end{aligned}$$

Sehingga nilai maksimum dari fungsi *likelihood* pada persamaan (2.20) diberikan oleh:

$$\begin{aligned}
 \max_{\hat{\Omega}} L(\boldsymbol{\beta}_\Omega, \sigma_\Omega^2) &= L(\hat{\Omega}) = L(\widehat{\boldsymbol{\beta}}_\Omega, \hat{\sigma}_\Omega^2) \\
 L(\hat{\Omega}) &= (2\pi\hat{\sigma}_\Omega^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\hat{\sigma}_\Omega^2} (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega) \right) \\
 &= (2\pi\hat{\sigma}_\Omega^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2} \frac{n(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)} \right) \\
 &= (2\pi\hat{\sigma}_\Omega^2)^{-\frac{n}{2}} \exp \left(-\frac{n}{2} \right) \\
 &= (2\pi\hat{\sigma}_\Omega^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \quad (2.23)
 \end{aligned}$$

Selanjutnya dengan memperhatikan parameter yang terdapat di bawah $H(\omega)$, diperoleh fungsi *likelihood* sebagai berikut:

$$\begin{aligned}
 L(\boldsymbol{\beta}_\omega, \sigma_\omega^2) &= (2\pi\sigma_\omega^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma_\omega^2} (\mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'_\omega\mathbf{X}[K]'\mathbf{y} + \boldsymbol{\beta}'_\omega\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta}_\omega) \right) \\
 &= -\frac{n}{2} \log(2\pi\sigma_\omega^2) - \frac{1}{2\sigma_\omega^2} (\mathbf{y}'\mathbf{y} - 2\sigma_\omega^2\boldsymbol{\beta}'_\omega\mathbf{X}[K]'\mathbf{y} + \sigma_\omega^2\boldsymbol{\beta}'_\omega\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta}_\omega) \quad (2.24)
 \end{aligned}$$

Estimasi parameter $\widehat{\beta}_\omega$ di bawah $H_0(\omega)$ diperoleh dengan menggunakan Fungsi *Lagrange Multiplier* (LM). Diberikan fungsi LM sebagai berikut:

$$F(\beta_\omega, \theta) = W(\beta_\omega) + 2\theta(c_j' \beta_\omega) \quad (2.25)$$

dengan

$$\begin{aligned} W(\beta_\omega) &= (\mathbf{y} - \mathbf{X}[K]\widehat{\beta}_\omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\beta}_\omega) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}[K]\beta_\omega - \beta_\omega'\mathbf{X}[K]'\mathbf{y} + \beta_\omega'\mathbf{X}[K]'\mathbf{X}[K]\beta_\omega \\ &= \mathbf{y}'\mathbf{y} - 2\beta_\omega'\mathbf{X}[K]'\mathbf{y} + \beta_\omega'\mathbf{X}[K]'\mathbf{X}[K]\beta_\omega \end{aligned}$$

Maka

$$F(\beta_\omega, \theta) = \mathbf{y}'\mathbf{y} - 2\beta_\omega'\mathbf{X}[K]'\mathbf{y} + \beta_\omega'\mathbf{X}[K]'\mathbf{X}[K]\beta_\omega + 2\theta(c_j' \beta_\omega) \quad (2.26)$$

Jika persamaan (2.26) diturunkan terhadap β_ω maka akan menghasilkan persamaan:

$$\begin{aligned} \frac{\partial F(\beta_\omega, \theta)}{\partial \beta_\omega} &= -2\mathbf{X}[K]'\mathbf{y} + 2\mathbf{X}[K]'\mathbf{X}[K]\beta_\omega + 2\theta c_j \\ &\quad -2\mathbf{X}[K]'\mathbf{y} + 2\mathbf{X}[K]'\mathbf{X}[K]\widehat{\beta}_\omega + 2\theta c_j = 0 \\ 2\mathbf{X}[K]'\mathbf{X}[K]\widehat{\beta}_\omega &= 2\mathbf{X}[K]'\mathbf{y} - 2\theta c_j \\ \widehat{\beta}_\omega &= (\mathbf{X}[K]'\mathbf{X}[K])^{-1}(\mathbf{X}[K]'\mathbf{y} - \theta c_j) \\ &= (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} - (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\theta c_j \\ &= \widehat{\beta}_\Omega - (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\theta c_j \end{aligned} \quad (2.27)$$

Jika persamaan (2.26) diturunkan terhadap θ maka akan menghasilkan persamaan:

$$\frac{\partial F(\beta_\omega, \theta)}{\partial \theta} = 2c_j' \beta_\omega$$

Jika persamaan di atas disamakan dengan nol, maka persamaan akan menjadi:

$$\begin{aligned} c_j' \widehat{\beta}_\omega &= 0 \\ c_j'(\widehat{\beta}_\Omega - (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\theta c_j) &= 0 \\ c_j' \widehat{\beta}_\Omega - c_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\theta c_j &= 0 \\ c_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\theta c_j &= c_j' \widehat{\beta}_\Omega \\ \theta &= (c_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}c_j)^{-1} c_j' \widehat{\beta}_\Omega \end{aligned} \quad (2.28)$$

Sehingga persamaan (2.27) akan menjadi

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\omega} &= \widehat{\boldsymbol{\beta}}_{\Omega} - (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \\ \widehat{\boldsymbol{\beta}}_{\omega} - \widehat{\boldsymbol{\beta}}_{\Omega} &= (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega}\end{aligned}\quad (2.29)$$

Untuk mendapatkan estimasi parameter σ_{ω}^2 , dilakukan dengan menurunkan persamaan (2.24) terhadap σ_{ω}^2 , akan didapatkan persamaan berikut ini.

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\beta}_{\omega}, \sigma_{\omega}^2)}{\partial \sigma_{\omega}^2} &= \frac{\partial}{\partial \sigma_{\omega}^2} (2\pi\sigma_{\omega}^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\omega}^2} \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'_{\omega}\mathbf{X}[K]'\mathbf{y} + \boldsymbol{\beta}'_{\omega}\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta}_{\omega}\right) \\ &= -\frac{n}{2}\left(\frac{2\pi}{2\pi\sigma_{\omega}^2}\right) + \frac{1}{2(\sigma_{\omega}^2)^2} (\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\boldsymbol{\beta}_{\omega}) \\ &\quad - \frac{n}{2}\left(\frac{2\pi}{2\pi\hat{\sigma}_{\omega}^2}\right) + \frac{1}{2(\hat{\sigma}_{\omega}^2)^2} (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}) = 0 \\ \frac{n}{2\hat{\sigma}_{\omega}^2} &= \frac{1}{2(\hat{\sigma}_{\omega}^2)^2} (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}) \\ \hat{\sigma}_{\omega}^2 &= \frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})}{n}\end{aligned}\quad (2.30)$$

Sehingga nilai maksimum dari fungsi *likelihood* pada persamaan (2.24) diberikan oleh:

$$\begin{aligned}\max_{\Omega} L(\boldsymbol{\beta}_{\omega}, \sigma_{\omega}^2) &= L(\hat{\omega}) = L(\widehat{\boldsymbol{\beta}}_{\omega}, \hat{\sigma}_{\omega}^2) \\ L(\hat{\omega}) &= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}_{\omega}^2} (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})\right) \\ &= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \frac{n(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})}\right) \\ &= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\right) \\ &= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}\end{aligned}\quad (2.31)$$

Selanjutnya diperoleh Rasio *Likelihood* sebagai berikut:

$$\begin{aligned}\lambda(x_1, \dots, x_h, y_i) &= \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{(2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}}{(2\pi\hat{\sigma}_{\Omega}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma}_{\omega}^2}{\hat{\sigma}_{\Omega}^2}\right)^{-\frac{n}{2}} \\ \lambda(x_1, \dots, x_h, y_i) &= \left(\frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})}{n} \frac{n}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})}\right)^{-\frac{n}{2}}\end{aligned}$$

$$\begin{aligned}
 &= \left(\frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})} \right)^{-\frac{n}{2}} \\
 &= \left(\frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})} \right)^{\frac{2}{n}} \\
 &= \left(\frac{D_1}{D} \right)^{\frac{2}{n}} \tag{2.32}
 \end{aligned}$$

dengan

$$D_1 = (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) \text{ dan } D = (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})$$

Selanjutnya persamaan D akan dijabarkan dengan cara mensubstitusikan $(-\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} + \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})$ pada persamaan D, sehingga akan diperoleh:

$$\begin{aligned}
 D &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} + \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} + \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}) \\
 &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} + \mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega}))'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega} + \mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})) \\
 &= ((\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})' + \mathbf{X}[K]'(\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})')((\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + \mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})) \\
 &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'\mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega}) + \\
 &\quad \mathbf{X}[K]'(\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + \mathbf{X}[K]'(\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})'\mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega}) \\
 &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + L(\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega}) + (\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})'M + \\
 &\quad (\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})'\mathbf{X}[K]'\mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega}) \tag{2.33}
 \end{aligned}$$

Nilai $\widehat{\boldsymbol{\beta}}_{\Omega}$ akan disubstitusikan pada ruas kedua dan ketiga pada persamaan (2.33)

yaitu $L = (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'\mathbf{X}[K]$ dan $M = \mathbf{X}[K]'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})$ sebagai berikut:

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'\mathbf{X}[K] \\
 &= (\mathbf{y} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y})'\mathbf{X}[K] \\
 &= (\mathbf{y}' - \mathbf{y}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]')\mathbf{X}[K] \\
 &= \mathbf{y}'\mathbf{X}[K] - \mathbf{y}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{X}[K] \\
 &= \mathbf{y}'\mathbf{X}[K] - \mathbf{y}'\mathbf{X}[K] \\
 &= 0
 \end{aligned}$$

Dan untuk M sebagai berikut:

$$\begin{aligned}
 \mathbf{M} &= \mathbf{X}[K]'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) \\
 &= \mathbf{X}[K]'(\mathbf{y} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y}) \\
 &= \mathbf{X}[K]'(\mathbf{y} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y}) \\
 &= \mathbf{X}[K]'\mathbf{y} - \mathbf{X}[K]'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \\
 &= \mathbf{X}[K]'\mathbf{y} - \mathbf{X}[K]'\mathbf{X}[K]\mathbf{y} \\
 &= 0
 \end{aligned}$$

Sehingga persamaan (2.33) akan menjadi:

$$D = (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + (\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})'\mathbf{X}[K]'\mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega}) \quad (2.34)$$

Jika

$$B = (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})$$

$$C = (\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})'\mathbf{X}[K]'\mathbf{X}[K](\widehat{\boldsymbol{\beta}}_{\Omega} - \widehat{\boldsymbol{\beta}}_{\omega})$$

maka persamaan (2.34) dapat ditulis menjadi: $D = B + C$. Selanjutnya persamaan C akan dijabarkan dengan mensubstitusikan persamaan (2.29) sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{\omega} - \widehat{\boldsymbol{\beta}}_{\Omega} = (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega}$$

maka akan diperoleh persamaan C sebagai berikut:

$$\begin{aligned}
 C &= \left((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \right)' \mathbf{X}[K]'\mathbf{X}[K] \\
 &\quad \left((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \right)
 \end{aligned}$$

Sehingga persamaan D menjadi:

$$\begin{aligned}
 D &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + \left((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \right)' \\
 &\quad \mathbf{X}[K]'\mathbf{X}[K] \left((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \right) \\
 &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + (\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{c}_j'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j' \\
 &\quad (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \\
 &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}) + (\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega})'(\mathbf{c}_j'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \\
 D &= D_1 + D_2 \tag{2.35}
 \end{aligned}$$

dengan

$$D_1 = (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)$$

$$D_2 = (\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega$$

Kemudian persamaan D akan disubstitusikan ke persamaan rasio *likelihood* yang telah diperoleh pada persamaan (2.32), sebagai berikut:

$$\begin{aligned} \lambda(x_1, \dots, x_h, y_i) &= \left(\frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\omega)} \right)^{\frac{2}{n}} = \left(\frac{D_1}{D} \right)^{\frac{2}{n}} \\ &= \left(\frac{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega) + (\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega} \right)^{\frac{2}{n}} \\ &= \left(\frac{1}{1 + \frac{(\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)}} \right)^{\frac{2}{n}} \\ &= \left(\frac{1}{1 + \frac{D_2}{D_1}} \right)^{\frac{2}{n}} \end{aligned} \quad (2.36)$$

Berdasarkan penjabaran diatas diperoleh suatu persamaan yang dinotasikan dengan Q sebagai berikut:

$$Q = \frac{D_2}{D_1} = \frac{(\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega}{(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)} \quad (2.37)$$

dimana

$$\begin{aligned} D_2 &= (\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j\widehat{\boldsymbol{\beta}}_\Omega \\ &= (\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y})'(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \\ &= \mathbf{y}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}'_j(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \\ &= \mathbf{y}'\mathbf{A}\mathbf{y} \end{aligned} \quad (2.38)$$

dengan

$$\mathbf{A} = \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'$$

Matrik \mathbf{A} adalah matriks yang simetris dan idempoten, maka dapat dinyatakan bahwa:

$$\frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\sigma^2} \sim \chi^2(r_1, \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/2\sigma^2) \quad (2.39)$$

Selanjutnya akan didapatkan nilai r_1 dan $\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/2\sigma^2$. Karena matriks \mathbf{A} adalah simetris dan idempoten, maka $r_1 = \text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A})$.

$$\begin{aligned} \text{tr}(\mathbf{A}) &= \text{tr}\left(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\right) \\ &= \text{tr}\left(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\right) \\ &= \text{tr}\left(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j\mathbf{I}(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\right) \\ &= \text{tr}(\mathbf{I}_{(1)}) \\ &= 1 \end{aligned}$$

Jadi, diperoleh $r_1 = \text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}) = 1$

Selanjutnya akan didapatkan nilai dari $\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/2\sigma^2$ dengan cara sebagai berikut:

$$\begin{aligned} \frac{\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}}{2\sigma^2} &= \frac{1}{2\sigma^2} ((\mathbf{X}[K]\boldsymbol{\beta})'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1} \\ &\quad \mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'(\mathbf{X}[K]\boldsymbol{\beta})) \end{aligned}$$

Akan dilakukan penjabaran pada pembilang sebagai berikut:

$$\begin{aligned} \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} &= (\mathbf{X}[K]\boldsymbol{\beta})'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1} \\ &\quad \mathbf{X}[K]'(\mathbf{X}[K]\boldsymbol{\beta}) \\ &= \mathbf{X}[K]'\boldsymbol{\beta}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1} \\ &\quad \mathbf{X}[K]'(\mathbf{X}[K]\boldsymbol{\beta}) \\ &= (\mathbf{c}_j'\boldsymbol{\beta})'(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'\boldsymbol{\beta} \\ &= 0 \end{aligned}$$

Sehingga dapat dinyatakan

$$\frac{D_2}{\sigma^2} \sim \chi^2(1)$$

Tahap kedua adalah dengan menjabarkan persamaan D_1 sebagai berikut:

$$\begin{aligned}
 D_1 &= (\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega) \\
 &= (\mathbf{y}' - \mathbf{X}[K]'\widehat{\boldsymbol{\beta}}_\Omega)'(\mathbf{y} - \mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega) \\
 &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega - \widehat{\boldsymbol{\beta}}_\Omega'\mathbf{X}[K]'\mathbf{y} + \widehat{\boldsymbol{\beta}}_\Omega'\mathbf{X}[K]'\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega \\
 &= \mathbf{y}'\mathbf{y} - 2\widehat{\boldsymbol{\beta}}_\Omega'\mathbf{X}[K]'\mathbf{y} + \widehat{\boldsymbol{\beta}}_\Omega'\mathbf{X}[K]'\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega
 \end{aligned}$$

Dengan mensubstitusikan nilai $\widehat{\boldsymbol{\beta}}_\Omega$ maka akan diperoleh:

$$\begin{aligned}
 D_1 &= \mathbf{y}'\mathbf{y} - 2\widehat{\boldsymbol{\beta}}_\Omega'\mathbf{X}[K]'\mathbf{y} + \widehat{\boldsymbol{\beta}}_\Omega'\mathbf{X}[K]'\mathbf{X}[K]\widehat{\boldsymbol{\beta}}_\Omega \\
 &= \mathbf{y}'\mathbf{y} - 2((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y})'\mathbf{X}[K]'\mathbf{y} + ((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y})' \\
 &\quad \mathbf{X}[K]'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \\
 &= \mathbf{y}'\mathbf{I}\mathbf{y} - 2\mathbf{y}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} + \mathbf{y}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \\
 &= \mathbf{y}'\mathbf{I}\mathbf{y} - \mathbf{y}'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y} \\
 &= \mathbf{y}'[\mathbf{I} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]']\mathbf{y} \\
 &= \mathbf{y}'\mathbf{B}\mathbf{y}
 \end{aligned} \tag{2.40}$$

dengan

$$\mathbf{B} = \mathbf{I} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'$$

Matrik \mathbf{B} adalah matriks yang simteris dan idempoten, maka dapat dinyatakan bahwa:

$$\frac{\mathbf{y}'\mathbf{B}\mathbf{y}}{\sigma^2} \sim \chi^2(r_1, \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}/2\sigma^2) \tag{2.41}$$

Selanjutnya akan didapatkan nilai r_2 dan $\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}/2\sigma^2$. Karena matrik \mathbf{B} adalah simetris dan idempoten, maka $r_1 = \text{rank}(\mathbf{B}) = \text{tr}(\mathbf{B})$.

$$\begin{aligned}
 \text{tr}(\mathbf{B}) &= \text{tr}(\mathbf{I} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]') \\
 &= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]') \\
 &= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_{(q+r)+1}) \\
 &= n - ((q+r)j + 1) \\
 &= n - (q+r)j - 1
 \end{aligned}$$

Jadi, diperoleh $r_2 = \text{rank}(\mathbf{B}) = \text{tr}(\mathbf{B}) = n - (q+r)j - 1$. Selanjutnya akan didapatkan nilai dari $\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}/2\sigma^2$ dengan cara sebagai berikut:

$$\begin{aligned}
 \frac{\boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}}{2\sigma^2} &= \frac{1}{2\sigma^2} \left((\mathbf{X}[K]\boldsymbol{\beta})'(\mathbf{I} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]')(\mathbf{X}[K]\boldsymbol{\beta}) \right) \\
 &= \frac{1}{2\sigma^2} \left(\mathbf{X}[K]'\boldsymbol{\beta}'\mathbf{X}[K]\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}[K]'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}(\mathbf{X}[K]\boldsymbol{\beta}) \right) \\
 &= \frac{1}{2\sigma^2} \left(\boldsymbol{\beta}'\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta} - (\boldsymbol{\beta}'\mathbf{X}[K]'\mathbf{X}[K]\boldsymbol{\beta}) \right) \\
 &= 0
 \end{aligned}$$

Sehingga dapat dinyatakan

$$\frac{D_1}{\sigma^2} \sim \chi^2(n - (q + r)j - 1)$$

Tahap selanjutnya adalah membuktikan bahwa D_2 dan D_1 independen. Pembuktian ini menggunakan Teorema 5.6b *Corollary 1* pada Rencher,dkk (2007). Sesuai dengan persamaan (2.38) dan persamaan (2.40) maka didapatkan

$$\mathbf{A} = \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'$$

$$\text{dan } \mathbf{B} = \mathbf{I} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'$$

sehingga

$$\begin{aligned}
 \mathbf{AB} &= \left(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right) \times \\
 &\quad \left(\mathbf{I} - \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right) \\
 &= \left(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right) - \\
 &\quad \left(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right. \\
 &\quad \left. \mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right) \\
 &= \left(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right) - \\
 &\quad \left(\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j(\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{c}_j)^{-1}\mathbf{c}_j'(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]' \right) \\
 &= 0
 \end{aligned}$$

Kajian pengujian hipotesis seperti yang diuraikan pada persamaan (2.20) sampai (2.41) telah dilakukan oleh Husni (2018). Selanjutnya, diperoleh bahwa $\mathbf{AB} = 0$ maka dapat dikatakan bahwa D_2 dan D_1 saling independen. Berdasarkan teorema tentang distribusi t dan distribusi F, serta D_2 dan D_1 terbukti saling independen maka

diperoleh statistik uji Q^* berdistribusi F dengan derajat bebas $n - (q + r)j - 1$, seperti berikut ini:

$$Q_F^* = \frac{\frac{D_2}{1}}{\frac{D_1}{n - (q + r)j - 1}} \sim F_{(1, n - (q + r)j - 1)} \quad (2.42)$$

dan diperoleh statistik uji Q^* berdistribusi t dengan derajat bebas $n - (q + r)j - 1$, seperti berikut ini:

$$\begin{aligned} Q_t^* &= \frac{\sqrt{D_2/1}}{\sqrt{D_1/n - (q + r)j - 1}} \sim t_{(n - (q + r)j - 1)} \\ &= \frac{Z^*}{\sqrt{D_1/n - (q + r)j - 1}} \sim t_{(n - (q + r)j - 1)} \end{aligned} \quad (2.43)$$

2.8 Pengujian Hipotesis Parameter Model Regresi Nonparametrik Spline *Truncated*

Pengujian hipotesis terhadap parameter dilakukan untuk mengetahui apakah suatu variabel memberikan pengaruh yang signifikan dalam model regresi atau tidak. Pada regresi nonparametrik spline, uji parameter akan dilakukan setelah mendapatkan model regresi dengan dengan titik knot optimal berdasarkan GCV yang paling minimum. Terdapat dua tahap pengujian parameter yaitu secara serentak dan secara parsial.

2.8.1 Pengujian Secara Serentak

Pengujian model secara serentak merupakan uji parameter kurva regresi secara simultan dengan menggunakan uji F . Misalkan diberikan suatu model regresi seperti pada persamaan (2.8). Maka hipotesis yang digunakan untuk menguji model secara serentak adalah sebagai berikut:

$$H_0 : \beta_0 = \beta_{1p} = \dots = \beta_{(q+r)j} = 0$$

H_1 : paling sedikit terdapat satu $\beta_0 \neq 0$ atau $\beta_{1p} \neq 0$ atau $\beta_{(q+r)j} \neq 0$; $l = 1, 2, \dots, q + r$

Selanjutnya diperoleh statistik uji F seperti pada persamaan (2.42) yang dapat dituliskan dalam bentuk sebagai berikut:

$$F_{hitung} = \frac{\frac{SSR}{(q+r)j}}{\frac{SSE}{n - (q+r)j - 1}} = \frac{MSR}{MSE} \quad (2.44)$$

dengan MSR (*Mean Square Regresi*) adalah hasil bagi antara jumlah kuadrat regresi dengan *df regresi*, sedangkan MSE (*Mean Square Error*) adalah hasil bagi dari jumlah kuadrat error dengan *df error* yang disajikan pada Tabel 2.1 menggunakan analisis ragam (ANOVA).

Tabel 2.1 Analisis of Varians (ANOVA)

Sumber	Derajat Bebas (df)	Jumlah Kuadrat (<i>Sum Square</i>)	Kuadrat Tengah (<i>Mean Square</i>)	F _{hitung}
Regresi	$(q+r)j$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{(q+r)j}$	$\frac{MSR}{MSE}$
Error	$n - (q+r)j - 1$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (q+r)j - 1}$	
Total	$n - 1$	$\sum_{i=1}^n (y_i - \bar{y})^2$	-	-

Daerah penolakan yang digunakan adalah tolak H_0 jika $F_{hitung} > F_{(\alpha, (q+r)j; n-(q+r)j-1)}$ atau $p - value < \alpha$. Apabila keputusan menolak H_0 maka dapat disimpulkan minimal terdapat satu parameter pada model regresi spline *truncated* yang signifikan terhadap model.

2.8.2 Pengujian Secara Parsial

Apabila dalam pengujian secara serentak didapatkan kesimpulan minimal terdapat satu parameter pada model regresi *spline truncated* yang signifikan, maka

perlu dilanjutkan uji secara parsial. Variabel prediktor dikatakan berpengaruh terhadap variabel respon apabila terdapat minimal satu parameter yang signifikan. Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \beta_l = 0$$

$$H_1 : \text{minimal ada } \beta_l \neq 0; l = 1, 2, \dots, q + r$$

Pengujian ini dilakukan dengan menggunakan uji t . Statistik uji yang digunakan adalah sebagai berikut :

$$t_{hitung} = \frac{\hat{\beta}_l}{SE(\hat{\beta}_l)} \quad (2.45)$$

dengan

$\hat{\beta}_l$: penaksir parameter β_l

$SE(\hat{\beta}_l)$: standart error dari $\hat{\beta}_l$

Nilai $SE(\hat{\beta}_l)$ didapatkan dari $\sqrt{var(\hat{\beta}_l)}$, $var(\hat{\beta}_l)$ merupakan elemen diagonal ke- l dari matriks sebagai berikut :

$$\begin{aligned} Var(\hat{\beta}) &= Var((\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{y}) \\ &= (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\text{Var}(\mathbf{y})(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]')' \\ &= (\mathbf{X}[K]'\mathbf{X}[K])^{-1}\sigma^2\mathbf{I}\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1} \\ &= \sigma^2(\mathbf{X}[K]'\mathbf{X}[K])^{-1}\mathbf{X}[K]'\mathbf{X}[K](\mathbf{X}[K]'\mathbf{X}[K])^{-1} \\ &= \sigma^2(\mathbf{X}[K]'\mathbf{X}[K])^{-1} \end{aligned}$$

Nilai σ^2 didekati dengan nilai MSE. Daerah penolakan yang digunakan adalah tolak H_0 jika nilai $|t_{hitung}| > t_{\alpha/2; (n-(q+r)j-1)}$ atau $p - value < \alpha$. Kesimpulan yang diperoleh jika tolak H_0 adalah parameter berpengaruh signifikan terhadap model (Drapper & Smith, 1992).

2.9 Kriteria Keباikan Model

Nilai koefisien determinasi (R^2) merupakan salah satu kriteria kebaikan model. Nilai R^2 menunjukkan seberapa besar model yang dihasilkan mampu menjelaskan

variabilitas data. Model yang baik adalah model yang memiliki nilai R^2 tinggi. Nilai R^2 diperoleh menggunakan persamaan (2.26).

$$R^2 = \frac{SSR}{SST} \times 100\% \quad (2.46)$$

Dalam pemilihan model juga akan memperhatikan banyak parameter yang digunakan pada model tersebut. Hal ini dijelaskan oleh prinsip parsimoni, dimana suatu model regresi yang baik adalah model regresi dengan banyak parameter sesedikit mungkin tetapi mempunyai R^2 yang cukup tinggi.

2.10 Pemeriksaan Asumsi Residual

2.10.1 Uji Identik

Pengujian ini dilakukan untuk mengetahui homogenitas varians residual model regresi. Apabila varians antar residual tidak homogen maka terjadi kasus heteroskedastisitas. Heteroskedastisitas dapat dideteksi dengan uji *Glejser*. Uji *Glejser* dilakukan dengan cara meregresikan nilai mutlak residual dengan variabel prediktornya.

$$|e_i| = \sum_{j=1}^h f_j(x_{ij}) + \varepsilon_i, \quad j = 1, 2, \dots, h$$

Hipotesis untuk uji *Glejser*, yaitu:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (residual identik)}$$

$$H_1 : \text{minimal terdapat satu } \sigma_i^2 \neq \sigma^2, \quad i = 1, 2, \dots, n \text{ (residual tidak identik).}$$

Statistik uji yang digunakan :

$$F_{hitung} = \frac{[\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}_i|)^2] / (v - 1)}{[\sum_{j=1}^n (|e_j| - |\hat{e}_j|)^2] / (n - v)} \quad (2.47)$$

Daerah penolakannya adalah tolak H_0 apabila $F_{hitung} > F_{\alpha, (v-1), (n-v)}$ dimana nilai v merupakan banyaknya parameter model *Glejser*. Jika H_0 ditolak maka dapat disimpulkan bahwa terdapat kasus heteroskedastisitas sehingga asumsi identik tidak terpenuhi (Gujarati, 2004).

2.10.2 Uji independen

Pengujian asumsi independen dilakukan untuk mengetahui apakah terdapat autokorelasi atau tidak pada residual. Ada beberapa cara untuk mendeteksi autokorelasi diantaranya adalah melihat plot ACF (*Autocorrelation Function*) dari residual. Untuk mendapatkan nilai ACF digunakan rumus sebagai berikut.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n-k} (e_t - \bar{e})(e_{t-k} - \bar{e})}{\sum_{t=1}^n (e_t - \bar{e})^2}, \quad k = 1, 2, 3, \dots \quad (2.48)$$

dimana, $\hat{\rho}_k$ merupakan korelasi antara e_t dan e_{t-k} , dan k adalah lag ke- k . Interval konfidensi $(1 - \alpha)100\%$ untuk autokorelasi $\hat{\rho}_k$ diberikan oleh persamaan (2.46).

$$-t_{\alpha/2, n-1} SE(\hat{\rho}_k) < \rho_k < t_{1-\alpha/2, n-1} SE(\hat{\rho}_k) \quad (2.49)$$

dengan

$$SE(\hat{\rho}_k) = \sqrt{\frac{1 + 2 \sum_{k=1}^{k-1} \hat{\rho}_k^2}{n}}$$

Apabila terdapat ACF ($\hat{\rho}_k$) yang keluar dari interval konfidensi, maka diindikasikan adanya kasus autokorelasi antar residual. Sebaliknya, jika semua nilai ACF berada di dalam batas interval maka tidak terdapat kasus autokorelasi atau asumsi independen terpenuhi (Wei, 2006).

2.10.3 Uji Normalitas

Tujuan uji normalitas terhadap residual adalah untuk mengetahui apakah dalam model regresi, residual mengikuti distribusi normal atau tidak. Jika residual tidak memenuhi asumsi normal maka pengujian parameter menjadi tidak akurat. Salah satu cara yang dapat digunakan untuk melakukan pengujian ini adalah uji *Kolmogorov-Smirnov* yang juga dikenal dengan uji kesesuaian model (*Goodness of fit test*). Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : F_0(x) = F(x)$ (Residual berdistribusi normal)

$H_1 : F_0(x) \neq F(x)$ (Residual tidak berdistribusi normal)

Statistik uji:

$$D = \text{maks}|F_0(x) - S_N(x)| \quad (2.50)$$

dengan $F_0(x)$ adalah fungsi peluang kumulatif distribusi normal sedangkan $S_N(x) = \frac{k}{n}$ adalah fungsi peluang kumulatif yang diobservasi dari satu sampel random dengan N observasi, k adalah banyaknya observasi yang sama atau kurang dari x , serta $F(x)$ adalah fungsi distribusi yang belum diketahui. Daerah penolakan pada uji *Kolmogorov Smirnov* yaitu tolak H_0 apabila $|D| > q_{(1-\alpha,n)}$ atau $-value < \alpha$, dimana nilai $q_{(1-\alpha,n)}$ didapatkan dari tabel *Kolmogorov Smirnov*. Keputusan yang didapatkan jika tolak H_0 adalah residual tidak berdistribusi normal (Daniel, 1989).

2.11 Teorema Dasar Terkait Dengan Aljabar Matriks

Beberapa teorema dasar terkait aljabar matriks yang digunakan untuk menyelesaikan estimasi parameter dan kajian pengujian hipotesis berikut ini berdasarkan Rencher dan Scaalje (2007).

1. Definisi matriks idempoten

Matrik A dikatakan idempoten jika $A^2 = A$.

2. Teorema 1

Jika A adalah $n \times p$ dan B adalah $p \times n$, maka $tr(AB) = tr(BA)$.

3. Teorema 2

Jika matrik A mempunyai *rank* r serta simetris dan *idempotent*, $rank(A) = tr(A) = r$.

4. Teorema 3

Jika $u = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$, dimana $\mathbf{a}' = (a_1, a_2, \dots, a_p)$ adalah vektor konstan. Maka

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}$$

5. Teorema 4

Jika $u = \mathbf{x}'\mathbf{A}\mathbf{x}$, dimana A matrik simetris konstan, maka

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

6. Teorema 5

Jika \mathbf{A} dan \mathbf{B} adalah matrik dengan ukuran $n \times m$, maka $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.

7. Teorema 6

Jika \mathbf{A} adalah matrik berukuran $n \times p$, dan \mathbf{B} adalah matrik berukuran $p \times n$ maka $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$.

8. Teorema 7

Jika $\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ dan \mathbf{A} matrik simetris dengan *rank* r , maka distribusi dari $\mathbf{y}'\mathbf{A}\mathbf{y}/\sigma^2$ adalah $\chi^2(r, \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/\sigma^2)$, jika dan hanya jika \mathbf{A} adalah *idempotent*.

9. Teorema 8

Jika $\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$, maka $\mathbf{y}'\mathbf{A}\mathbf{y}$ dan $\mathbf{y}'\mathbf{B}\mathbf{y}$ adalah *idempotent* jika dan hanya jika $\mathbf{A}\mathbf{B} = 0$ (atau ekuivalen, $\mathbf{B}\mathbf{A} = 0$).

10. Distribusi F

Jika U_1 dan U_2 peubah yang saling bebas dan berdistribusi *chi-square* dengan derajat kebebasan masing-masing p_1 dan p_2 maka statistic:

$$F = \frac{U_1/p_1}{U_2/p_2}$$

11. Distribusi *t-student*

Jika Z adalah variabel random yang berdistribusi normal standar $N(0,1)$ dan U adalah variabel random yang mengikuti distribusi *chi-square* dengan derajat bebas p yaitu $\chi^2(p)$, dengan Z dan U saling independen, maka statistik:

$$T = \frac{Z}{\sqrt{U/p}} \sim t_{(p)}$$

BAB III METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Badan Pusat Statistik (BPS) tahun 2020 yang bersumber dari data SAKERNAS (Survei Angkatan Kerja Nasional). Objek dalam penelitian ini adalah Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2020. Variabel yang digunakan dalam penelitian ini terbagi dalam dua variabel yaitu, variabel respon (y) dan variabel prediktor (x). Variabel yang menjadi perhatian dalam penelitian ini adalah tingkat pengangguran terbuka (TPT).

3.2 Deskripsi Variabel

Variabel yang digunakan dalam penelitian ini terdiri dari dua variabel yaitu variabel respon dan variabel prediktor.

1. Tingkat Pengangguran Terbuka (y)

Tingkat pengangguran terbuka merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja. Pengangguran adalah bagian dari angkatan kerja yang sekarang ini tidak bekerja dan sedang aktif mencari pekerjaan atau dapat diartikan sebagai keadaan pengangguran terbuka. Tingkat Pengangguran Terbuka (TPT) dapat dinyatakan dalam rumus sebagai berikut.

$$\text{Tingkat Pengangguran} = \frac{\text{jumlah pengangguran}}{\text{jumlah angkatan kerja}} \times 100\%$$

2. Persentase penduduk miskin(x_1)

Persentase penduduk miskin merupakan persentase penduduk yang berada dibawah garis kemiskinan. Rumus presentase penduduk miskin ditulis sebagai berikut.

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

Keterangan:

$\alpha = 0$

z = Garis kemiskinan

y_i = Rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ($i = 1, 2, 3, \dots, q$), $y_i < z$

q = Banyaknya penduduk yang berada dibawah garis kemiskinan

n = Jumlah penduduk

3. Rata-rata Lama Sekolah (x_2)

Rata-rata lama sekolah merupakan jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal. Rumus Rata-rata lama sekolah dapat ditulis sebagai berikut.

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

Keterangan:

RLS = Rata-rata lama sekolah penduduk usia 25 tahun ke atas

x_i = Lama sekolah penduduk ke- i yang berusia 25 tahun

n = Jumlah penduduk usia 25 tahun ke atas

4. *Dependency Ratio* (x_3)

Dependency Ratio merupakan perbandingan antara jumlah penduduk umur 0-14 tahun, 22 ditambah dengan jumlah penduduk 65 tahun ke atas (keduanya disebut dengan bukan angkatan kerja) dibandingkan dengan jumlah penduduk usia 15-64 tahun (angkatan kerja).

Jika jumlah pengangguran tinggi maka rasio ketergantungan tinggi pula dikarenakan negara memiliki tanggungan yang besar untuk penduduk dimana

kondisi tersebut mampu menghambat pembangunan dan masalah sosial lainnya. Rumus rasio ketergantungan dapat ditulis sebagai berikut.

$$DR = \frac{P_{(0-14)} + P_{(65+)}}{P_{(15-64)}} \times 100$$

Keterangan :

DR = Rasio Ketergantungan

$P_{(0-14)}$ = Jumlah penduduk usia muda

$P_{(65+)}$ = Jumlah penduduk usia tua

$P_{(15-64)}$ = Jumlah penduduk usia produktif

5. Indeks Pembangunan Manusia (x_4)

Indeks pembangunan manusia merupakan pengukuran perbandingan dari harapan hidup, pendidikan, dan standar hidup.

Struktur data yang digunakan dalam penelitian ini dijabarkan seperti dalam Tabel 3.1.

Tabel 3.1 Struktur Data

Kabupaten/Kota	y	x_1	x_2	x_3	x_4
1	y_1	x_{11}	x_{21}	x_{31}	x_{41}
2	y_2	x_{12}	x_{22}	x_{32}	x_{42}
3	y_3	x_{13}	x_{23}	x_{33}	x_{43}
⋮	⋮	⋮	⋮	⋮	⋮
24	y_{124}	x_{124}	x_{224}	x_{324}	x_{424}

3.3 Langkah Penelitian

Sebelum melakukan tahap penelitian, terlebih dahulu dilakukan tahap *Pre-Processing* data yang akan diolah sebagai berikut:

1. Menyiapkan data angka pengangguran dari data BPS Provinsi Sulawesi Selatan.
2. Menyimpan data untuk variabel prediktor (x_1 , x_2 , x_3 , dan x_4) per kabupaten/kota di Provinsi Sulawesi Selatan tahun 2020.

Selanjutnya metode dan tahapan penelitian yang akan dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut:

1. Melakukan eksplorasi data.
2. Membuat *scatter plot* antara tingkat pengangguran terbuka dengan masing-masing variabel prediktor untuk mengetahui pola hubungan yang terjadi.
3. Memodelkan variabel respon menggunakan model regresi nonparametrik *spline truncated* dengan 1-3 titik knot.
4. Memilih titik knot optimal berdasarkan nilai *Generalized Cross Validation* (GCV) yang paling minimum.
5. Melakukan pengujian signifikansi parameter model regresi nonparametrik *spline truncated* secara serentak dan parsial sesuai dengan persamaan 2.44 dan persamaan 2.45.
6. Melakukan uji asumsi residual identik, independen, dan berdistribusi normal (IIDN) berdasarkan model regresi nonparametrik *spline truncated*.
7. Menguji kebaikan model dengan menghitung koefisien determinasi (R^2) sesuai dengan persamaan 2.46.
8. Melakukan interpretasi model.

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai metode spline *truncated* terhadap tingkat pengangguran terbuka di Provinsi Sulawesi Selatan. Penelitian ini menggunakan data dari Badan Pusat Statistik Provinsi Sulawesi Selatan (BPS) tahun 2020 yang bersumber dari data SAKERNAS (Survei Angkatan Kerja Nasional).

Objek pengamatan adalah 24 kabupaten/kota yang akan diduga angka penganggurannya berdasarkan variabel yang telah ditetapkan. Variabel respon yang digunakan dalam penelitian ini adalah Tingkat Pengangguran Terbuka (TPT) sedangkan variabel prediktornya adalah persentase penduduk miskin (x_1), rata-rata lama sekolah (x_2), *dependency ratio* (x_3), dan Indeks Pembangunan Manusia (x_4).

4.1 Eksplorasi Data

Sebelum melakukan pemodelan dan pengujian hipotesis, terlebih dahulu dilakukan eksplorasi data. Hasil dari eksplorasi data dapat digunakan sebagai informasi awal untuk analisis tahap selanjutnya.

Eksplorasi data TPT mencakup besarnya nilai rata-rata, standar deviasi, nilai minimum, dan maksimum, serta variansi. Besarnya nilai ringkasan statistik dapat dilihat dalam Tabel 4.1.

Tabel 4.1 Statistik Deskriptif Keseluruhan Data

Variabel	Rata-rata	Standar Deviasi	Minimum	Maksimum
y	6,15	6,07	2,31	31
x_1	9,418	2,919	4,54	14,58
x_2	8,148	1,176	6,59	11,21
x_3	53,97	6,99	40,84	71,38
x_4	70,485	4,028	64,26	82,25

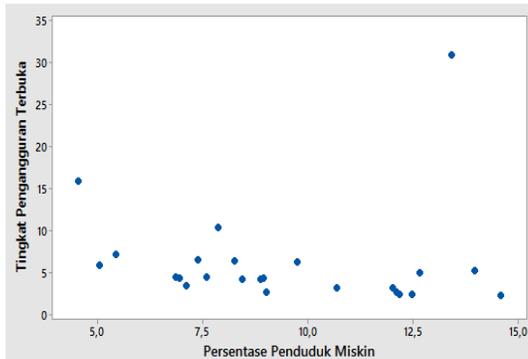
Keterangan : y = Tingkat Pengangguran Terbuka, x_1 = Persentase Penduduk Miskin, x_2 = *Dependency Ratio*, x_3 = Rata-rata Lama Sekolah, x_4 = Indeks Pembangunan Manusia

Pada Tabel 4.1 diketahui bahwa rata-rata tingkat pengangguran terbuka di Provinsi Sulawesi Selatan tahun 2020 adalah 6,15 persen dengan standar deviasi yaitu 6,07 persen. Daerah yang memiliki tingkat pengangguran terbuka tertinggi adalah Kabupaten Luwu Utara sebesar 31 persen dan daerah dengan tingkat pengangguran terbuka terkecil adalah Kabupaten Jeneponto yaitu sebesar 2,31 persen. Adapun karakteristik dari masing-masing variabel prediktor, sebagai berikut:

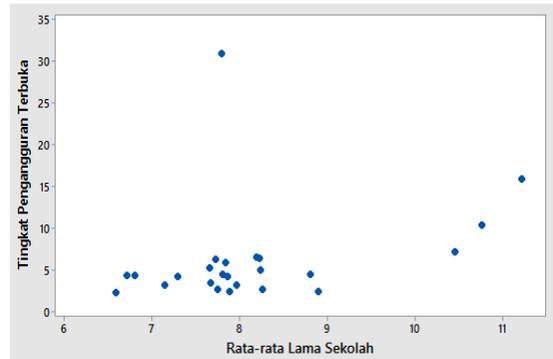
- a. Rata-rata persentase penduduk miskin di Sulawesi Selatan tahun 2020 adalah 9,418 dengan standar deviasi sebesar 2,919. Kabupaten yang memiliki persentase penduduk miskin di Sulawesi Selatan yaitu Kota Makassar sebesar 4,54 sedangkan persentase penduduk miskin yang paling tinggi terjadi di Kabupaten Jeneponto sebesar 14,58.
- b. Rata-rata lama sekolah di Sulawesi Selatan tahun 2020 memiliki rata-rata sebesar 8,148 tahun dengan standar deviasi sebesar 1,176. Kota/Kabupaten yang memiliki rata-rata lama sekolah terendah di Sulawesi Selatan yaitu Kabupaten Jeneponto sebesar 6,59 tahun sedangkan rata-rata lama sekolah yang paling tinggi terjadi di Kota Makassar sebesar 11,21 tahun.
- c. Rata-rata *dependency ratio* di Sulawesi Selatan tahun 2020 adalah 53,97 persen dengan standar deviasi sebesar 6,99 persen. Kabupaten yang memiliki *dependency ratio* terendah di Sulawesi Selatan yaitu Kota Makassar sebesar 40,84 persen sedangkan *dependency ratio* yang paling tinggi terjadi di Kabupaten Toraja Utara sebesar 71,38 persen.
- d. Rata-rata indeks pembangunan manusia di Sulawesi Selatan tahun 2020 adalah 70,485 dengan standar deviasi sebesar 4,028. Kabupaten yang memiliki indeks pembangunan manusia terendah di Sulawesi Selatan yaitu Kabupaten Jeneponto sebesar 64,26 sedangkan indeks pembangunan manusia yang paling tinggi terjadi di Kota Makassar sebesar 82,25.

Eksplorasi data selanjutnya adalah membuat *scatter plot*. Hal ini bertujuan untuk mengidentifikasi bentuk pola hubungan antara variabel respon dengan masing-

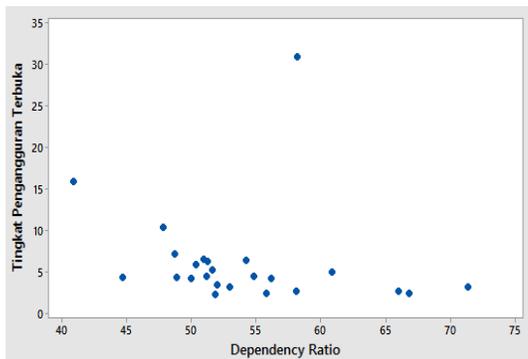
masing variabel prediktor. *Scatter plot* antara variabel respon dan variabel prediktor yang diduga mempengaruhinya. Berikut ini hasil identifikasi mengenai pola hubungan antara Tingkat Pengangguran Terbuka (TPT) dengan variabel yang diduga berpengaruh yaitu persentase penduduk miskin (x_1), rata-rata lama sekolah (x_2), *dependency ratio* (x_3), dan indeks pembangunan manusia (x_4).



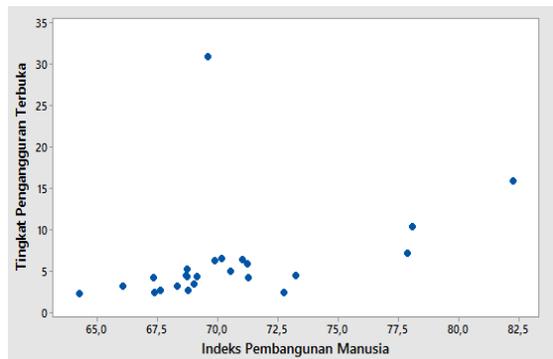
Scatter plot (y) dengan (x₁)



Scatter plot (y) dengan (x₂)



Scatter plot (y) dengan (x₃)



Scatter plot (y) dengan (x₄)

Gambar 4.1 *Scatter plot* antara variabel respon dan variabel prediktor pada data tingkat pengangguran terbuka tahun 2020.

Berdasarkan Gambar 4.1 terlihat bahwa kurva antara TPT terhadap masing-masing variabel prediktor tidak menunjukkan adanya kecenderungan pola atau tidak diketahui bentuk kurva regresinya. Kurva regresi yang diidentifikasi melalui plot-plot yang tersebar menunjukkan pola yang tidak beraturan atau dapat dikatakan tidak mengikuti suatu pola tertentu. Maka diperlukan pendekatan nonparametrik yang

merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu, salah satunya dengan metode spline *truncated*.

4.2 Pemodelan TPT Provinsi Sulawesi Selatan Menggunakan Regresi Nonparametrik Spline *Truncated*

Model regresi nonparametrik spline *truncated* digunakan setelah mengetahui bentuk pola hubungan antara Tingkat Pengangguran Terbuka (TPT) dengan variabel-variabel yang diduga berpengaruh. Secara umum, model regresi nonparametrik *spline truncated* berorde q dengan titik knot K_1, K_2, \dots, K_r dapat dinyatakan sebagai berikut.

$$y_i = \beta_{0j} + \sum_{l=1}^q \beta_{jl} x_{ji}^l + \sum_{h=1}^r \beta_{j(q+h)} (x_{ji} - K_{jh})_+^q ; j = 1, 2, \dots, p$$

4.3 Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama untuk melihat perubahan perilaku pada data. Untuk memperoleh model spline *truncated* yang optimal diperlukan strategi yang tepat dalam memilih dan menentukan lokasi knot agar tidak terjadi *overfitting*. Metode yang digunakan untuk mencari titik knot optimal adalah *Generalized Cross Validation* (GCV). Titik knot optimal diperoleh dari nilai GCV yang paling minimum.

4.3.1 Pemilihan Titik Knot Optimal Spline *Truncated* Linear dengan Satu Titik Knot

Pemilihan titik knot yang optimal diawali dengan menggunakan satu titik knot pada masing-masing variabel prediktor. Model regresi nonparametrik spline *truncated* dengan menggunakan satu titik knot dan 4 variabel prediktor adalah sebagai berikut

$$\hat{y} = \beta_0 + \beta_{11}x_{1i} + \beta_{12}(x_{1i} - K_{11})_+^1 + \beta_{21}x_{2i} + \beta_{22}(x_{2i} - K_{21})_+^1 + \beta_{31}x_{3i} + \beta_{32}(x_{3i} - K_{31})_+^1 + \beta_{41}x_{4i} + \beta_{42}(x_{4i} - K_{41})_+^1$$

Nilai GCV yang dihasilkan dengan menggunakan regresi nonparametrik spline *truncated* dengan satu knot disajikan dalam Tabel 4.2.

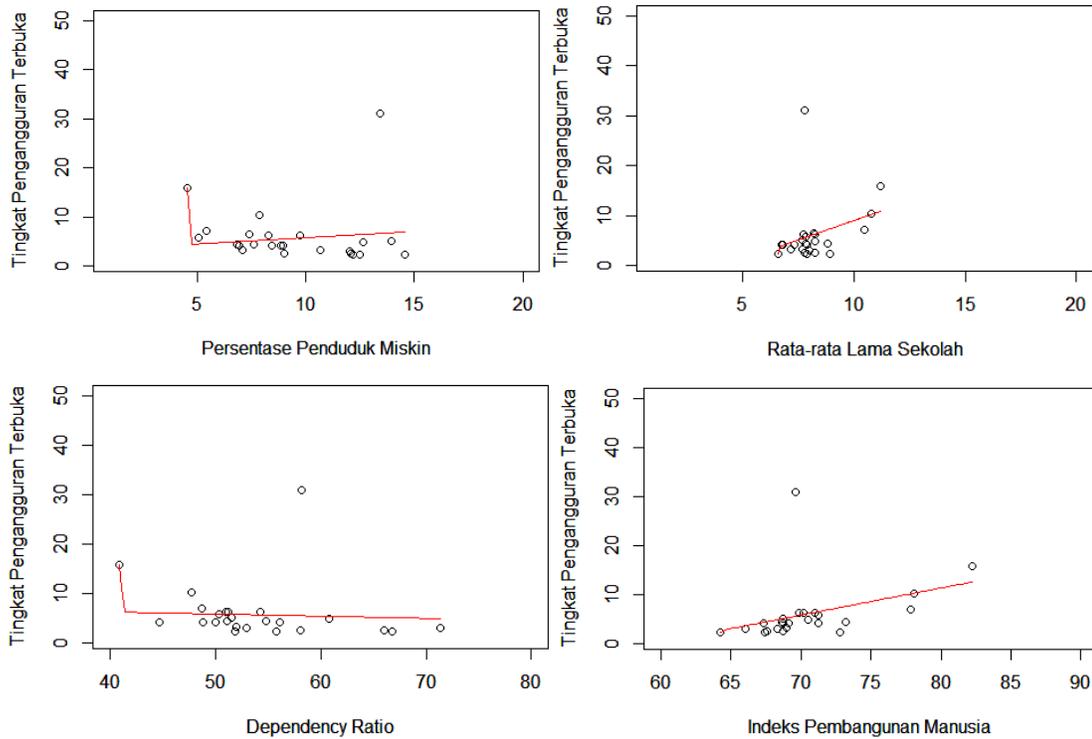
Tabel 4. 2 Nilai GCV dengan Satu Titik Knot Spline *Truncated* Linear

No.	Knot				GCV
	x_1	x_2	x_3	x_4	
1.	4,74	6,68	41,46	64,63	48,94
2.	13,76	10,83	68,89	80,78	53,24
3.	10,28	9,23	58,29	74,54	53,64
4.	4,95	6,78	42,09	64,99	53,72
5.	10,48	9,32	58,91	74,91	54,65
6.	10,07	9,14	57,67	74,17	55,16
7.	14,38	11,12	70,76	81,88	55,24
8.	14,17	11,02	70,13	81,52	55,24
9.	13,97	10,93	69,51	81,15	55,24
10.	10,69	9,42	59,54	75,27	55,80

Berdasarkan Tabel 4.2 dapat diketahui bahwa nilai GCV minimum yang diperoleh dengan satu titik knot yaitu sebesar 48,94 dimana titik knot optimumnya pada masing-masing variabel prediktor yaitu sebagai berikut.

$$K_{11} = 4,74 ; \quad K_{21} = 6,68 ; \quad K_{31} = 41,46 ; \quad K_{41} = 64,63$$

Selanjutnya kurva estimasi model spline linear dengan satu titik knot yang disajikan pada Gambar 4.2.



Gambar 4.2 Estimasi Kurva Spline Linear Satu Titik Knot

4.3.2 Pemilihan Titik Knot Optimal Spline *Truncated* Linear dengan Dua Titik Knot

Model regresi nonparametrik spline *truncated* dengan menggunakan dua titik knot dan 4 variabel prediktor adalah sebagai berikut

$$\hat{y} = \beta_0 + \beta_{11}x_{1i} + \beta_{12}(x_{1i} - K_{11})_+^1 + \beta_{13}(x_{1i} - K_{12})_+^1 + \beta_{21}x_{2i} + \beta_{22}(x_{2i} - K_{21})_+^1 + \beta_{23}(x_{2i} - K_{22})_+^1 + \beta_{31}x_{3i} + \beta_{32}(x_{3i} - K_{31})_+^1 + \beta_{33}(x_{3i} - K_{32})_+^1 + \beta_{41}x_{4i} + \beta_{42}(x_{4i} - K_{41})_+^1 + \beta_{43}(x_{4i} - K_{42})_+^1$$

Nilai GCV yang dihasilkan dengan menggunakan regresi nonparametrik spline *truncated* dengan dua knot disajikan dalam Tabel 4.3.

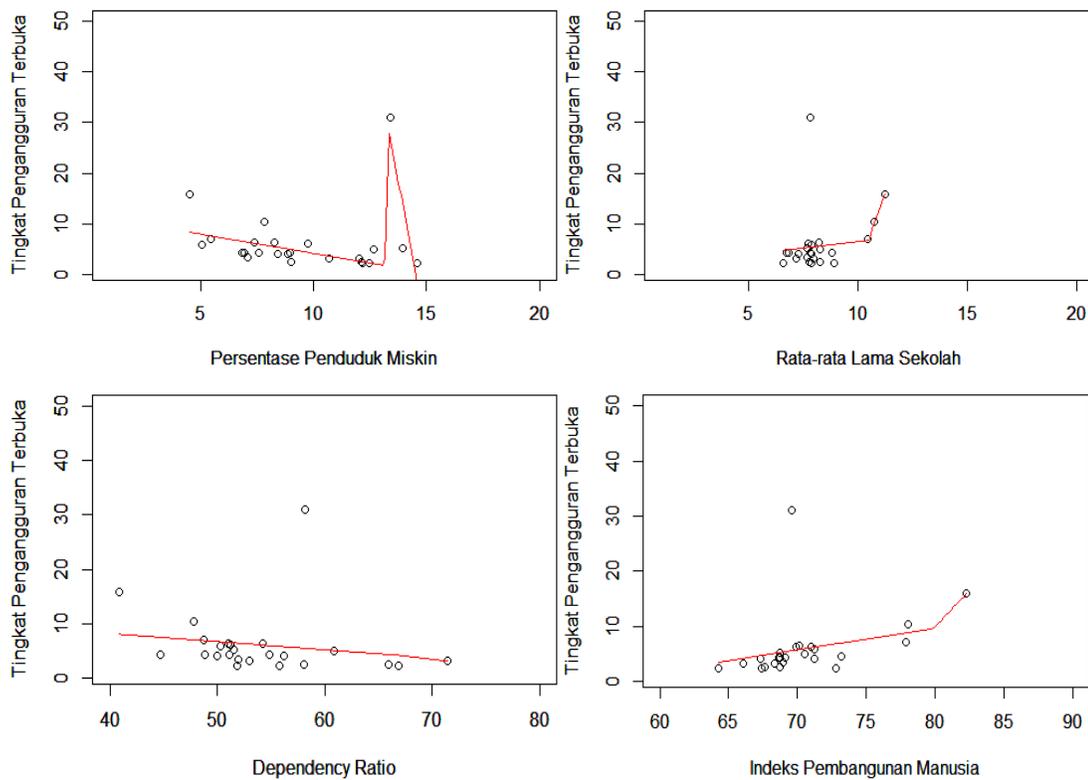
Tabel 4. 3 Nilai GCV dengan Dua Titik Knot Spline *Truncated* Linear

No.	Knot				GCV
	x_1	x_2	x_3	x_4	
1.	13,15	10,55	67,02	79,68	15,28
	13,35	10,64	67,64	80,05	
2.	12,94	10,46	66,39	79,31	17,26
	13,35	10,64	67,64	80,05	
3.	12,94	10,46	66,39	79,31	17,26
	13,15	10,55	67,02	79,68	
4.	12,74	10,36	65,77	78,95	18,69
	12,94	10,46	66,39	79,31	
5.	12,53	10,27	65,15	78,58	18,83
	13,35	10,64	67,64	80,05	
6.	12,74	10,36	65,77	78,94	18,93
	13,15	10,55	67,02	79,68	
7.	12,74	10,36	65,77	78,95	18,93
	13,35	10,64	67,64	80,05	
8.	12,53	10,27	65,15	78,58	20,33
	13,15	10,55	67,02	79,68	
9.	12,33	10,17	64,52	78,21	22,80
	13,35	10,64	67,64	80,05	
10.	12,53	10,27	65,15	78,58	26,46
	12,94	10,46	66,39	79,31	

Berdasarkan Tabel 4.3 dapat diketahui bahwa nilai GCV minimum yang diperoleh dengan dua titik knot yaitu sebesar 15,28, dimana titik knot optimumnya pada masing-masing variabel prediktor yaitu sebagai berikut.

$$\begin{aligned}
 (K_{11} = 13,15 & \quad ; K_{12} = 13,35) \\
 (K_{21} = 10,55 & \quad ; K_{22} = 10,64) \\
 (K_{31} = 67,02 & \quad ; K_{32} = 67,64) \\
 (K_{41} = 79,68 & \quad ; K_{42} = 80,05)
 \end{aligned}$$

Selanjutnya kurva estimasi model spline linear dengan dua titik knot yang disajikan pada Gambar 4.3.



Gambar 4.3 Estimasi Kurva Spline Linear Dua Titik Knot

4.3.3 Pemilihan Titik Knot Optimal Spline *Truncated* Linear dengan Tiga Titik Knot

Setelah mendapatkan knot optimum dengan nilai GCV minimum dari satu titik knot dan dua titik knot, maka tahap selanjutnya adalah pemilihan titik knot optimum dengan menggunakan tiga titik knot dengan cara yang serupa. Berikut merupakan model spline *truncated* linear dengan tiga titik knot pada kasus Tingkat Pengangguran Terbuka di Sulawesi Selatan tahun 2020.

$$\hat{y} = \beta_0 + \beta_{11}x_{1i} + \beta_{12}(x_{1i} - K_{11})_+^1 + \beta_{13}(x_{1i} - K_{12})_+^1 + \beta_{14}(x_{1i} - K_{13})_+^1 + \beta_{21}x_{2i} + \beta_{22}(x_{2i} - K_{21})_+^1 + \beta_{23}(x_{2i} - K_{22})_+^1 + \beta_{24}(x_{2i} - K_{23})_+^1 + \beta_{31}x_{3i} + \beta_{32}(x_{3i} - K_{31})_+^1 + \beta_{33}(x_{3i} - K_{32})_+^1 + \beta_{34}(x_{3i} - K_{33})_+^1 + \beta_{41}x_{4i} + \beta_{42}(x_{4i} - K_{41})_+^1 + \beta_{43}(x_{4i} - K_{42})_+^1 + \beta_{44}(x_{4i} - K_{43})_+^1$$

Nilai GCV dari pemodelan dengan menggunakan tiga titik knot ditunjukkan oleh Tabel 4.4.

Tabel 4. 4 Nilai GCV dengan Tiga Titik Knot Spline *Truncated* Linear

No	Knot				GCV
	x_1	x_2	x_3	x_4	
1.	4,74	6,68	41,46	64,63	2,16
	13,15	10,55	67,02	79,68	
	13,35	10,64	67,64	80,05	
2.	4.74	6.68	41.46	64.63	2,90
	12.53	10.27	65.15	78.58	
	13.56	10.74	68.26	80.41	
3.	4.74	6.68	41.46	64.63	2,95
	12.53	10.27	65.15	78.58	
	13.76	10.83	68.89	80.78	
4.	4.74	6.68	41.46	64.63	3,34
	12.53	10.27	65.15	78.58	

	13.35	10.64	67.64	80.05	
5.	4.95	6.78	42.09	64.99	3,45
	12.53	10.27	65.15	78.58	
	13.56	10.74	68.26	80.41	
6.	4.95	6.78	42.09	64.99	3,51
	12.53	10.27	65.15	78.58	
	13.76	10.83	68.89	80.78	
7.	4.95	6.78	42.09	64.99	3,98
	12.53	10.27	65.15	78.58	
	13.35	10.64	67.64	80.05	
8.	5.15	6.87	42.71	65.36	4,05
	12.53	10.27	65.15	78.58	
	13.76	10.83	68.89	80.78	
9.	5.15	6.87	42.71	65.36	4,29
	12.53	10.27	65.15	78.58	
	13.56	10.74	68.26	80.41	
10.	5.15	6.87	42.71	65.36	4,96
	12.53	10.27	65.15	78.58	
	13.35	10.64	67.64	80.05	

Berdasarkan Tabel 4.4 dapat diketahui bahwa nilai GCV minimum yang diperoleh dengan tiga titik knot yaitu sebesar 2,16 dimana titik knot optimumnya pada masing-masing variabel prediktor yaitu sebagai berikut.

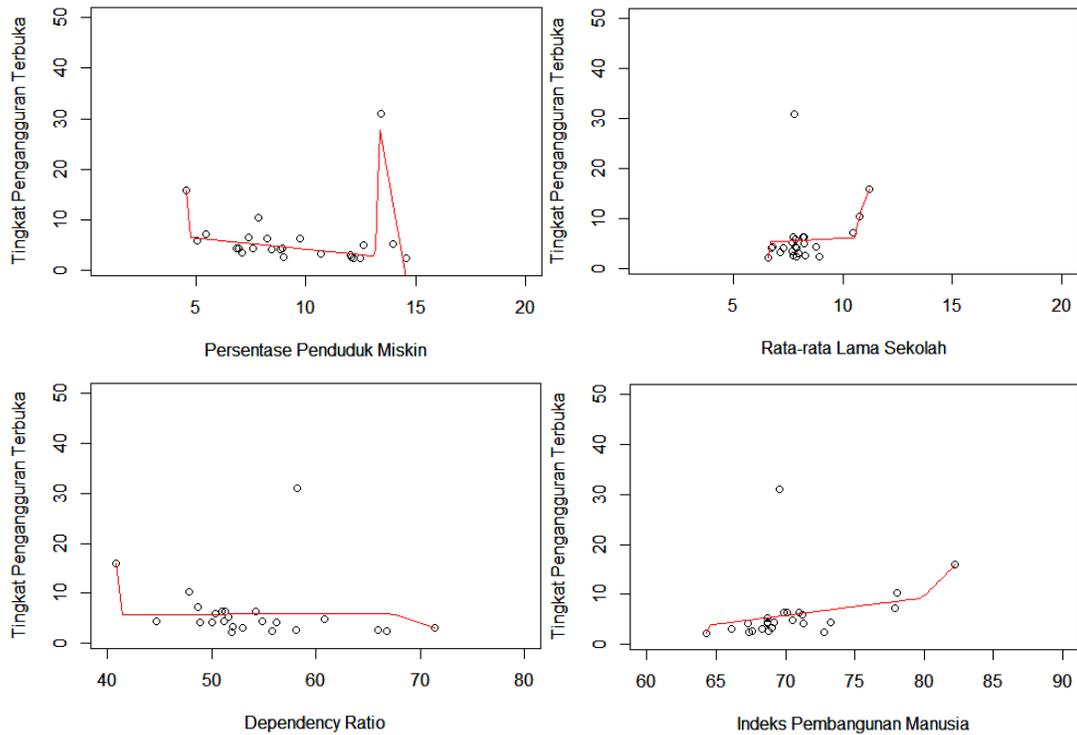
$$(K_{11} = 4,74 ; \quad K_{12} = 13,15 ; \quad K_{13} = 13,35)$$

$$(K_{21} = 6,68 ; \quad K_{22} = 10,55 ; \quad K_{23} = 10,64)$$

$$(K_{31} = 41,46 ; \quad K_{32} = 67,02 ; \quad K_{33} = 67,64)$$

$$(K_{41} = 64,63 ; \quad K_{42} = 79,68 ; \quad K_{43} = 80,05)$$

Selanjutnya kurva estimasi model spline linear dengan tiga titik knot yang disajikan pada Gambar 4.4.



Gambar 4.4 Estimasi Kurva Spline Linear Tiga Titik Knot

Berdasarkan hasil pemilihan titik knot yang paling optimum dengan menggunakan satu knot, dua knot, dan titik knot maka berikut adalah ringkasan GCV minimum yang dihasilkan.

Tabel 4.5 Perbandingan Nilai GCV Minimum

Model	Nilai GCV
Model 1 titik knot	48,94
Model 2 titik knot	15,28
Model 3 titik knot	2,16

Berdasarkan Tabel 4.5 terlihat bahwa model yang memiliki GCV minimum yaitu model spline *truncated* linear dengan menggunakan tiga titik knot dengan nilai GCV sebesar 2,16. Maka model nonparametrik spline *truncated* paling terbaik dengan menggunakan tiga titik knot pada variabel-variabel yang mempengaruhi tingkat pengangguran terbuka di Sulawesi Selatan adalah sebagai berikut.

$$\hat{y} = \beta_0 + \beta_{11}x_{1i} + \beta_{12}(x_{1i} - K_{11})_+^1 + \beta_{13}(x_{1i} - K_{12})_+^1 + \beta_{14}(x_{1i} - K_{13})_+^1 + \beta_{21}x_{2i} + \beta_{22}(x_{2i} - K_{21})_+^1 + \beta_{23}(x_{2i} - K_{22})_+^1 + \beta_{24}(x_{2i} - K_{23})_+^1 + \beta_{31}x_{3i} + \beta_{32}(x_{3i} - K_{31})_+^1 + \beta_{33}(x_{3i} - K_{32})_+^1 + \beta_{34}(x_{3i} - K_{33})_+^1 + \beta_{41}x_{4i} + \beta_{42}(x_{4i} - K_{41})_+^1 + \beta_{43}(x_{4i} - K_{42})_+^1 + \beta_{44}(x_{4i} - K_{43})_+^1$$

4.4 Penaksiran Parameter Tingkat Pengangguran Terbuka di Provinsi Sulawesi Selatan Tahun 2020

Model nonparametrik Spline *Truncated* terbaik dihasilkan melalui perolehan titik knot optimum. Berdasarkan proses pemilihan titik knot yang telah dilakukan sebelumnya diketahui bahwa titik knot paling optimum adalah dengan menggunakan tiga titik knot. Hasil estimasi parameter dengan menggunakan tiga titik knot adalah sebagai berikut.

Tabel 4.6 Estimasi Model Spline Linier Tiga Titik Knot

Parameter	Estimator
β_0	5,36
β_{11}	12,13
β_{12}	-12,08
β_{13}	146,79
β_{14}	-195,26
β_{21}	-43,55
β_{22}	43,78
β_{23}	7,60
β_{24}	5,39
β_{31}	109,19
β_{32}	-109,35
β_{33}	0,25

β_{34}	0,21
β_{41}	-66,45
β_{42}	66,71
β_{43}	15,29
β_{44}	13,10

Hasil estimasi parameter pada Tabel 4.6 membentuk persamaan model regresi nonparametrik spline *truncated* berikut.

$$\hat{y} = 5,36 + 12,13x_{1i} - 12,08(x_{1i} - 4,74)_+^1 + 146,79(x_{1i} - 13,15)_+^1 - 195,26(x_{2i} - 13,35)_+^1 - 43,55x_{2i} + 43,78(x_{2i} - 6,68)_+^1 + 7,60(x_{2i} - 10,55)_+^1 + 5,39(x_{2i} - 10,64)_+^1 + 109,19x_{3i} - 109,35(x_{3i} - 41,46)_+^1 + 0,25(x_{3i} - 67,02)_+^1 + 0,21(x_{3i} - 67,64)_+^1 - 66,45x_{4i} + 66,71(x_{4i} - 64,63)_+^1 + 15,29(x_{4i} - 79,68)_+^1 + 13,10(x_{4i} - 80,05)_+^1$$

4.5 Pengujian Parameter Model Regresi

Setelah mendapatkan model terbaik dengan nilai GCV terkecil maka dilanjutkan melakukan pengujian signifikansi parameter untuk mengetahui variabel prediktor apa saja yang berpengaruh signifikan terhadap model.

4.5.1 Pengujian Parameter Secara Serentak

Pada uji serentak akan diketahui apakah variabel-variabel prediktor yang digunakan berpengaruh secara serentak terhadap model. Berikut hasil ANOVA untuk model regresi nonparametrik spline *truncated*.

Tabel 4.7 ANOVA Model Regresi Spline *Truncated*

Sumber Variasi	Df	Sum of Square (SS)	Mean Square (MS)	F_{hitung}
Regresi	16	831,8639	51,9915	23,97387
Error	7	15,18072	2,168674	
Total	23	847,0447		

Tabel 4.7 menunjukkan nilai $F_{hitung} = (23,97) > F_{(0,05,16,7)} = (3,49)$ menghasilkan kesimpulan tolak H_0 . Maka minimal ada satu parameter yang signifikan terhadap tingkat pengangguran terbuka di Provinsi Sulawesi Selatan tahun 2020. Selanjutnya dilakukan pengujian secara parsial untuk melihat parameter yang berpengaruh secara signifikan.

4.5.2 Pengujian Parameter Secara Parsial

Untuk melakukan pengujian signifikansi parameter secara parsial dilakukan uji sebagai berikut:

Tabel 4.8 Uji Parsial Model Regresi Nonparametrik Spline *Truncated*

Variabel	Parameter	Estimator	t_{hitung}	P-value	Keputusan
Konstan	β_0	5,36	7,09	0,000195	Signifikan
x_1	β_{11}	12,13	7,01	0,000211	Signifikan
	β_{12}	-12,08	-7,24	0,000172	Signifikan
	β_{13}	146,79	15,83	0,974559	Signifikan
	β_{14}	-195,26	-16,25	0,815478	Signifikan
x_2	β_{21}	-43,55	-7,11	0,000193	Signifikan
	β_{22}	43,78	7,05	0,000203	Signifikan
	β_{23}	7,60	0,89	0,402813	Tidak Signifikan
	β_{24}	5,39	1,18	0,275132	Tidak Signifikan
x_3	β_{31}	109,19	7,11	0,000192	Signifikan
	β_{32}	-109,35	-7,11	0,000193	Signifikan
	β_{33}	0,25	0,92	0,387067	Tidak Signifikan
	β_{34}	0,21	0,92	0,387067	Tidak Signifikan
x_4	β_{41}	-66,45	-7,10	0,000194	Signifikan
	β_{42}	66,71	7,11	0,000191	Signifikan
	β_{43}	15,29	5,26	0,001175	Signifikan
	β_{44}	13,10	5,26	0,001175	Signifikan

Tabel 4.8 menunjukkan parameter yang signifikan dan tidak signifikan dalam setiap variabel yang disertai nilai t_{hitung} dan $p-value$. Apabila membandingkan antara $p-value$ dengan taraf signifikansi sebesar 0,05 maka didapatkan 13 parameter

yang memberikan pengaruh secara signifikan terhadap model. Parameter yang signifikan tersebut merupakan parameter pada variabel persentase penduduk miskin (x_1), rata-rata lama sekolah (x_2), *dependency ratio* (x_3), dan indeks pembangunan manusia (x_4). Keempat variabel tersebut dikatakan sebagai variabel yang berpengaruh secara signifikan terhadap TPT di Sulawesi Selatan tahun 2020.

4.6 Pengujian Asumsi Residual

Pengujian asumsi residual dilakukan untuk mengetahui apakah residual yang dihasilkan dari model telah memenuhi asumsi identik, independen, dan berdistribusi normal (IIDN) atau tidak. Apabila suatu model regresi tidak memenuhi asumsi residual IIDN, namun memiliki kriteria model terbaik dan parameter signifikan maka model tersebut tetap tidak layak digunakan untuk memodelkan tingkat pengangguran terbuka di Sulawesi Selatan tahun 2020.

4.6.1 Asumsi Identik

Pengujian asumsi residual identik digunakan untuk mengetahui apakah terjadi kasus heteroskedastisitas atau tidak, dengan kata lain variansi residual dari model harus homogen. Pengujian asumsi residual identik dilakukan menggunakan uji Glejser dengan hipotesis sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (residual identik)}$$

$$H_1 : \text{minimal terdapat satu } \sigma_i^2 \neq \sigma^2, i = 1, 2, \dots, n \text{ (residual tidak identik)}$$

Hasil uji Glejser disajikan dalam Tabel 4.9 sebagai berikut.

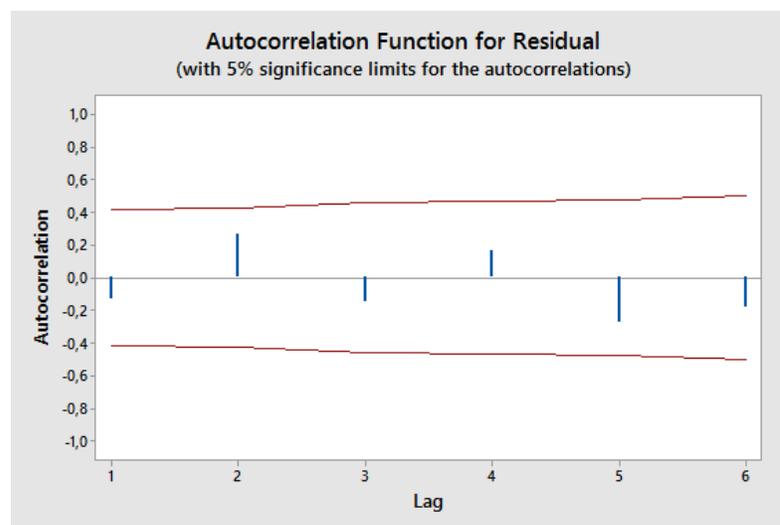
Tabel 4.9 ANOVA dari Uji Glejser

Sumber	Derajat Bebas (df)	Jumlah Kuadrat (<i>Sum Square</i>)	Kuadrat Tengah (<i>Mean Square</i>)	F _{hitung}
Regresi	16	3,245275	0,2028297	0,3539788
Error	7	4,010997	0,5729996	
Total	23	7,256273		

Berdasarkan Tabel 4.9, dengan menggunakan taraf signifikan sebesar 5% maka diperoleh keputusan terima H_0 karena nilai $F_{hitung} = 0,35 < F_{(0,05;16,7)} = 3,49$. Sehingga dapat disimpulkan bahwa tidak terjadi kasus heteroskedastisitas atau dengan kata lain variansi antar residual homogen. Hal ini berarti asumsi residual identik telah terpenuhi.

4.6.2 Asumsi Independen

Residual yang independen adalah residual yang tidak terjadi autokorelasi antar residualnya. Pemeriksaan asumsi independen pada penelitian ini menggunakan plot *Autocorrelation Function* (ACF). Apabila ada nilai autokorelasi yang keluar dari batas atas maupun batas bawah interval konfidensi, maka dapat disimpulkan terdapat kasus autokorelasi antar residual. Gambar 4.7 menunjukkan bahwa tidak terlihat adanya nilai autokorelasi yang keluar batas interval konfidensi (garis merah), sehingga dapat disimpulkan bahwa tidak terdapat autokorelasi antar residual. Hal ini berarti asumsi residual independen telah terpenuhi.



Gambar 4.5 ACF Residual

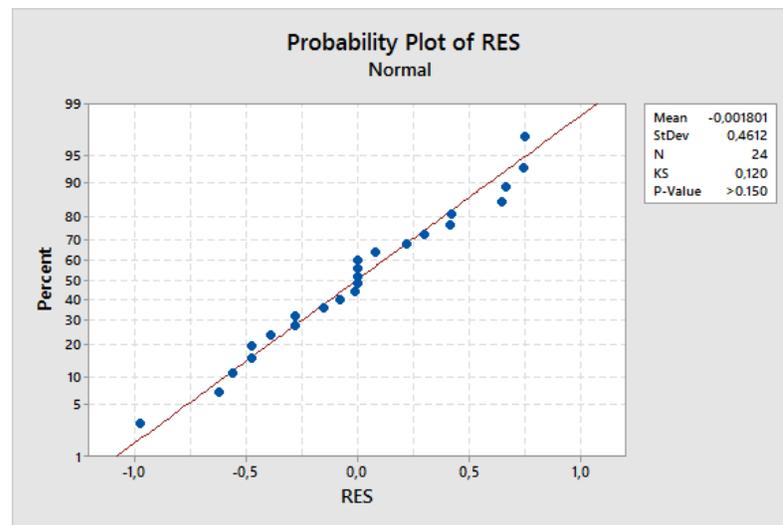
4.6.3 Asumsi Distribusi Normal

Pengujian asumsi residual berdistribusi normal dilakukan untuk mengetahui apakah residual telah berdistribusi normal atau tidak. Pengujian ini dilakukan menggunakan uji *Kolmogorov Smirnov* dengan hipotesis sebagai berikut.

$$H_0 : F(x) = F_0(x)$$

$$H_1 : F(x) \neq F_0(x)$$

Hasil uji *Kolmogorov Smirnov* disajikan pada Gambar 4.8 sebagai berikut.



Gambar 4.6 Uji Normalitas

Berdasarkan gambar 4.3 nilai *P-value* > 0,150 yang berarti lebih dari $\alpha(0,05)$ sehingga menjadikan kesimpulan H_0 ditolak. Hal tersebut menjadikan residual memenuhi asumsi mengikuti distribusi normal.

4.6.4 Koefisien Determinasi

Nilai koefisien determinasi (R^2) menunjukkan seberapa besar kebaikan model regresi dalam menjelaskan variabilitas tingkat pengangguran terbuka di Sulawesi Selatan.

$$R^2 = \frac{SS_{regresi}}{SS_{total}} \times 100\%$$

$$\begin{aligned}
&= \frac{831,8639}{847,0447} \times 100\% \\
&= 98,21\%
\end{aligned}$$

Berdasarkan perhitungan didapatkan nilai R^2 sebesar 98,21 persen. Hal ini berarti model regresi nonparametrik spline *truncated* yang didapatkan mampu menjelaskan variabilitas tingkat pengangguran terbuka di Sulawesi Selatan sebesar 98,21 persen, sedangkan sisanya dijelaskan oleh variabel lain yang tidak masuk dalam model.

4.6.5 Interpretasi Model Regresi Nonparametrik Spline *Truncated*

Setelah melakukan pengujian parameter dan semua asumsi residual terpenuhi, selanjutnya adalah menginterpretasikan model regresi yang telah didapatkan. Model yang terbentuk menggunakan titik knot optimum ditunjukkan pada persamaan sebagai berikut.

$$\begin{aligned}
\hat{y} = & 5,36 - 12,13x_{1i} - 12,08(x_{1i} - 4,74)_+^1 + 146,79(x_{1i} - 13,15)_+^1 - \\
& 195,26(x_{1i} - 13,35)_+^1 - 43,55x_{2i} + 43,78(x_{2i} - 6,68)_+^1 + 7,60(x_{2i} - 10,55)_+^1 + \\
& 5,39(x_{2i} - 10,64)_+^1 + 109,19x_{3i} - 109,35(x_{3i} - 41,46)_+^1 + 0,25(x_{3i} - 67,02)_+^1 \\
& + 0,21(x_{3i} - 67,64)_+^1 - 66,45x_{4i} + 66,71(x_{4i} - 64,63)_+^1 + 15,29(x_{4i} - 79,68)_+^1 + \\
& 13,10(x_{4i} - 80,05)_+^1
\end{aligned}$$

Berdasarkan model tersebut, maka dapat diinterpretasikan dari masing-masing variabel yang berpengaruh terhadap tingkat pengangguran terbuka di Sulawesi Selatan tahun 2020 sebagai berikut.

1. Dengan mengasumsikan variabel x_2 , x_3 dan x_4 konstan, maka pengaruh variabel persentase penduduk miskin (x_1) terhadap tingkat pengangguran terbuka di Sulawesi Selatan dapat ditulis sebagai berikut.

$$\hat{y} = 12,13x_{1i} - 12,08(x_{1i} - 4,74)_+^1 + 146,79(x_{1i} - 13,15)_+^1 - 195,26(x_{1i} - 13,35)_+^1$$

$$\hat{y} = \begin{cases} 12,13x_{1i} & ; x_1 < 4,74 \\ 57,26 + 0,05x_1 & ; 4,74 \leq x_1 < 13,15 \\ -1873,03 + 146,84x_1 & ; 13,15 \leq x_1 < 13,35 \\ 733,69 - 48,42x_1 & ; x_1 \geq 13,35 \end{cases}$$

Variabel persentase penduduk miskin berpengaruh secara signifikan terhadap tingkat pengangguran terbuka di Sulawesi Selatan. Berdasarkan model tersebut dapat diketahui bahwa apabila faktor tingkat pengangguran terbuka dengan persentase penduduk miskin kurang dari 4,74% maka setiap kenaikan 1% persentase penduduk miskin cenderung meningkatkan TPT sebesar 12,13%. Apabila persentase penduduk miskin berada diantara 4,74% hingga 13,15% maka setiap kenaikan 1% persentase penduduk miskin cenderung meningkatkan TPT sebesar 0,05%. Apabila persentase penduduk miskin diantara 13,15% hingga 13,35% maka setiap kenaikan 1% persentase penduduk miskin cenderung meningkatkan TPT sebesar 146,84 persen. Selanjutnya apabila persentase penduduk miskin lebih dari 13,35% maka setiap kenaikan 1% persentase penduduk miskin cenderung menurunkan TPT sebesar 48,42 persen. Koefisien bernilai positif artinya terjadi hubungan positif antara persentase penduduk miskin dengan TPT yang mengindikasikan bahwa apabila persentase penduduk miskin meningkat maka TPT juga mengalami peningkatan.

2. Dengan mengasumsikan variabel x_1 , x_3 dan x_4 konstan, maka pengaruh variabel rata-rata lama sekolah (x_2) terhadap tingkat pengangguran terbuka di Sulawesi Selatan dapat ditulis sebagai berikut.

$$\hat{y} = -43,55x_{2i} + 43,78(x_{2i} - 6,68)_{+}^{\frac{1}{2}} + 7,60(x_{2i} - 10,55)_{+}^{\frac{1}{2}} + 5,39(x_{2i} - 10,64)_{+}^{\frac{1}{2}}$$

$$\hat{y} = \begin{cases} -43,55x_{2i} & ; x_2 < 6,68 \\ -292,45 + 0,23x_2 & ; 6,68 \leq x_2 < 10,55 \\ -372,63 + 7,83x_2 & ; 10,55 \leq x_2 < 10,64 \\ -429,98 + 13,22x_2 & ; x_2 \geq 10,64 \end{cases}$$

Variabel rata-rata lama sekolah berpengaruh secara signifikan terhadap tingkat pengangguran terbuka di Sulawesi Selatan. Berdasarkan model tersebut dapat diketahui bahwa apabila faktor tingkat pengangguran terbuka dengan rata-rata lama sekolah kurang dari 6,68% maka setiap kenaikan 1% rata-rata lama sekolah cenderung menurunkan TPT sebesar 43,55%. Apabila rata-rata lama sekolah berada diantara 6,68% hingga 10,55% maka setiap kenaikan 1% rata-rata lama sekolah cenderung meningkatkan TPT sebesar 0,23%. Apabila rata-rata lama sekolah diantara 10,55% hingga 10,64% maka setiap kenaikan 1% rata-rata lama sekolah cenderung meningkatkan TPT sebesar 7,83%. Dan apabila rata-rata lama sekolah lebih dari 10,64% maka setiap kenaikan 1% rata-rata lama sekolah cenderung meningkatkan TPT sebesar 13,22 persen. Dalam hal ini diketahui bahwa terjadi hubungan positif antara rata-rata lama sekolah dengan TPT mengindikasikan bahwa apabila rata-rata lama sekolah meningkat, maka TPT juga akan meningkat. Semakin lama masyarakat menempuh dan lulus tamatan pendidikan formal, maka semakin tinggi pula kemampuan dan kesempatan masyarakat tersebut untuk bekerja sehingga dapat menurunkan TPT.

3. Dengan mengasumsikan variabel x_1 , x_3 dan x_4 konstan, maka pengaruh variabel *dependency ratio* (x_3) terhadap tingkat pengangguran terbuka di Sulawesi Selatan dapat ditulis sebagai berikut.

$$\hat{y} = 109,19x_{3i} - 109,35(x_{3i} - 41,46)_+^1 + 0,25(x_{3i} - 67,02)_+^1 + 0,21(x_{3i} - 67,64)_+^1$$

$$\hat{y} = \begin{cases} 109,19x_{3i} & ; \quad x_3 < 41,46 \\ 67,89 - 0,16x_3 & ; \quad 41,46 \leq x_3 < 67,02 \\ 51,14 + 0,09x_3 & ; \quad 67,02 \leq x_3 < 67,64 \\ 36,93 + 0,30x_3 & ; \quad x_3 \geq 67,64 \end{cases}$$

Berdasarkan hasil pengujian, menunjukkan variabel *dependency ratio* berpengaruh secara signifikan terhadap tingkat pengangguran terbuka di Sulawesi Selatan. Dari model tersebut dapat diketahui bahwa apabila faktor tingkat pengangguran terbuka dengan *dependency ratio* kurang dari 41,46%

maka setiap kenaikan 1% *dependency ratio* cenderung meningkatkan TPT sebesar 109,19%. Apabila *dependency ratio* berada diantara 41,46% hingga 67,02% maka setiap kenaikan 1% *dependency ratio* cenderung menurunkan TPT sebesar 0,16%. Apabila *dependency ratio* berada diantara 67,02% hingga 67,64 persen maka setiap kenaikan 1% *dependency ratio* cenderung meningkatkan TPT sebesar 0,09%. Dan apabila *dependency ratio* lebih dari 67,64% maka setiap kenaikan 1% *dependency ratio* cenderung meningkatkan TPT sebesar 0,30%. Dalam hal ini diketahui bahwa terjadi hubungan positif antara *dependency ratio* dengan TPT di Sulawesi Selatan mengindikasikan semakin tingginya persentase *dependency ratio* maka jumlah pengangguran juga semakin tinggi dikarenakan negara memiliki tanggungan yang besar untuk penduduk dimana kondisi tersebut mampu menghambat pembangunan dan masalah sosial lainnya.

4. Dengan mengasumsikan variabel x_1 , x_2 dan x_3 konstan, maka pengaruh variabel indeks pembangunan manusia (x_4) terhadap tingkat pengangguran terbuka di Sulawesi Selatan dapat ditulis sebagai berikut.

$$\hat{y} = -66,45x_{4i} + 66,71(x_{4i} - 64,63)_+^1 + 15,29(x_{4i} - 79,68)_+^1 + 13,10(x_{4i} - 80,05)_+^1$$

$$\hat{y} = \begin{cases} -66,45x_{4i} & ; x_4 < 64,63 \\ -4311,47 + 0,26x_4 & ; 64,63 \leq x_4 < 79,68 \\ -5529,77 + 15,55x_4 & ; 79,68 \leq x_4 < 80,05 \\ -6578,43 + 28,65x_4 & ; x_4 \geq 80,05 \end{cases}$$

Berdasarkan hasil pengujian, menunjukkan variabel indeks pembangunan manusia (IPM) berpengaruh secara signifikan terhadap tingkat pengangguran terbuka di Sulawesi Selatan. Dari model tersebut dapat diketahui bahwa apabila faktor tingkat pengangguran terbuka dengan IPM kurang dari 64,63% maka setiap kenaikan 1% IPM cenderung menurunkan TPT sebesar 66,45%. Apabila IPM diantara 64,63% hingga 79,68% maka setiap kenaikan 1% IPM cenderung

meningkatkan TPT sebesar 0,26%. Apabila IPM diantara 79,68% hingga 80,05% maka setiap kenaikan 1% IPM cenderung meningkatkan TPT sebesar 15,55%. Dan apabila IPM lebih dari 80,05% maka setiap kenaikan 1% IPM cenderung meningkatkan TPT sebesar 28,65%. Dalam hal ini diketahui bahwa terjadi hubungan positif antara IPM dengan TPT di Sulawesi Selatan mengindikasikan bahwa apabila angka IPM di suatu wilayah meningkat maka akan menyebabkan meningkatnya TPT. Temuan hasil regresi pada variabel ini bertentangan dengan Burhanuddin (2015) yang menyimpulkan bahwa ketika angka IPM meningkat maka akan menyebabkan menurunnya TPT. Namun sesuai dengan penelitian Nurcholis (2014) yang mendapatkan hasil bahwa angka IPM berpengaruh positif terhadap TPT. Hal tersebut menunjukkan bahwa pada kondisi tertentu, hubungan antara indeks pembangunan manusia bisa positif maupun negatif terhadap TPT.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan pada bab sebelumnya maka dapat ditarik kesimpulan sebagai berikut.

1. Model regresi nonparametrik spline *truncated* terbaik merupakan model spline *truncated* linear dengan menggunakan tiga titik knot. Model regresi nonparametrik yang dihasilkan adalah sebagai berikut.

$$\begin{aligned} \hat{y} = & 5,36 + 12,13x_{1i} - 12,08(x_{1i} - 4,74)_+^1 + 146,79(x_{1i} - 13,15)_+^1 - \\ & 195,26(x_{1i} - 13,35)_+^1 - 43,55x_{2i} + 43,78(x_{2i} - 6,68)_+^1 + \\ & 7,60(x_{2i} - 10,55)_+^1 + 5,39(x_{2i} - 10,64)_+^1 + 109,19x_{3i} - \\ & 109,35(x_{3i} - 41,46)_+^1 + 0,25(x_{3i} - 67,02)_+^1 + 0,21(x_{3i} - 67,64)_+^1 - \\ & 66,45x_{4i} + 66,71(x_{4i} - 64,63)_+^1 + 15,29(x_{4i} - 79,68)_+^1 + \\ & 13,10(x_{4i} - 80,05)_+^1 \end{aligned}$$

Hasil estimasi model menunjukkan bahwa persentase penduduk miskin, rata-rata lama sekolah, *dependency ratio*, dan indeks pembangunan manusia di Sulawesi Selatan tahun 2020 menunjukkan indikator yang positif artinya tingkat pengangguran terbuka di Sulawesi Selatan mengalami kenaikan dan hal ini diimbangi dengan meningkatnya persentase penduduk miskin, rata-rata lama sekolah, *dependency ratio*, dan indeks pembangunan manusia.

2. Berdasarkan pengujian hipotesis, diperoleh hasil untuk pengujian parameter secara serentak dengan uji F nilai $F_{hitung} = 23,97 > F_{(0,05,16,7)} = 3,49$ menghasilkan kesimpulan tolak H_0 . Maka minimal ada satu parameter yang signifikan terhadap tingkat pengangguran terbuka di Provinsi Sulawesi Selatan tahun 2020. Kemudian untuk pengujian parameter secara parsial didapatkan bahwa keempat variabel berpengaruh secara signifikan terhadap tingkat pengangguran terbuka di Sulawesi Selatan tahun 2020 diantaranya yaitu

variabel persentase penduduk miskin (x_1), rata-rata lama sekolah (x_2), *dependency ratio* (x_3), dan indeks pembangunan manusia (x_4). Nilai koefisien determinasi didapatkan yaitu 98,21% sehingga dapat dikatakan bahwa model regresi nonparametrik spline *truncated* yang dihasilkan merupakan model yang baik dan layak untuk memodelkan TPT di Sulawesi Selatan.

5.2 Saran

Penelitian selanjutnya dapat mengembangkan regresi nonparametrik spline *truncated* dengan *spatial*, untuk mengetahui pengaruh perbedaan antar wilayah. Sehingga input terhadap pemerintah mengenai kebijakan suatu wilayah akan lebih tepat akan faktor-faktor apa saja yang berpengaruh.

DAFTAR PUSTAKA

- Amalia, E., & Sari, L. K. (2019). Analisis Spasial Untuk Mengidentifikasi Tingkat Pengangguran Terbuka Berdasarkan Kabupaten/Kota di Pulau Jawa Tahun 2017. *Indonesian Journal of Statistic and Its Applications (eISSN:2599-0802) Vol 3 No 3 (2019)*, 202-215.
- Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developments and their applications. *Statistics Survey Vol. 1*, 16–55.
- Antoniadis, A. G. (1994). Methods for Curve Estimation. *JASA Vol.89, No.428, 1994*, 1340-1353.
- Arjun, D. A. (2019). Analisis Faktor-faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Kalimantan Menggunakan Regresi Nonparametrik Spline Truncated. *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya 2019 Terbitan I, 5 Oktober 2019*, 115-121.
- BPS. (2009). *Berita Resmi Statistik Keadaan Ketenagakerjaan Jawa Barat*. Jawa Barat: Badan Pusat Statistik Provinsi Jawa Barat.
- Budiantara. (2005). "Regresi Spline Linier", *Makalah Pembicara Utama pada Seminar Nasional Matematika*. Semarang: FMIPA Universitas Diponegoro.
- Budiantara, I. (2006). Model Spline Dengan Knots Optimal. *Jurnal Ilmu Dasar, FMIPA Universitas Jember, Vol. 7*, 77-85.
- Budiantara, I. (2009). *Spline Dalam Regresi Nonparametrik dan Semiparametrik : Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Budiantara, I. (2011). Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter. *Prosiding Seminar Nasional MIPA Undiksha, Undiksha, pp, 9-28*.
- Burhanuddin, M. (2015). *Pengaruh Produk Domestik Regional Bruto (PDRB), Upah Minimum Kab/Kota (UMK), Indeks Pembangunan Manusia (IPM) Terhadap Pengangguran Terbuka di Provinsi Banten Periode 2008-2013*. UIN Jakarta.

- Casella, G., & Berger, R. (2001). *Statistical Inference, 2nd edition*. California: Duxbury Press.
- Daniel, W. W. (1989). *Statistika Non Parametrik*. Jakarta: Gramedia.
- Drapper, N., & dan Smith, H. (1992). *Analisis Regresi Terapan. Edisi Kedua*. Jakarta: PT.Gramedia Pustaka Utama.
- Eubank. (1999). Faktor-faktor Yang Mempengaruhi Angka Gizi Bruk di Jawa Timur Dengan Pendekatan Regresi Nonparametrik Spline. *Jurnal Sains dan Seni ITS, Vol. 1, No,1*.
- Eubank, R. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Ferdiana, K. (2017). *Pengujian Hipotesis Simultan dalam Regresi Semiparametrik Spline Truncated (Studi Kasus: Angka Partisipasi KAsar SLTA Tahun 2015 di Provinsi Jawa Timur*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Griggs, W. (2013). *Penalized Spline Regression and its Applications*. United States: Whitman College.
- Gujarati, D. (2004). *Basic Econometrics*. New York: Mc Gwra Hill,Inc,640-645.
- Gusti, O. (2011). *Regresi Semiparametrik Spline Dalam Memodelkan Hasil*. Malang: Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Hardle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.
- Husni, I. (2018). *Pengujian Hipotesis Parsial untuk Parameter Model Regresi Nonparametrik Spline Truncated Multivariabel (Aplikasi: Data Kematian Ibu Provinsi Nusa Tenggara Timur)*. Surabaya: Masters thesis, Institut Teknologi Sepuluh.
- Indahwati, Sadik K, & Nurmasari R. (2008). *Pendekatan Metode Pemulusan Kernel pada Pendugaan Area Kecil*. Yogyakarta: Universitas Negeri Yogyakarta.
- Jhingan, M. (2014). *Ekonomi Pembangunan dan Perencanaan*. Jakarta: Raja Grafindo persada.

- Kurnia, A. (2008). *Modifikasi General Regression dan Pendekatan Nonparametrik Pada Pendugaan Area Kecil*. Bogor: Institut Pertanian Bogor.
- Lyche, T. d. (2008). *Spline Methods Draft*. Departmen of Informatics for Applications University of Oslo.
- Mankiw, N. G. (2012). *Pegantar Ekonomi Makro*. Jakarta: Salemba Empat.
- Prasaja, M. H. (2011). Pengantar Investasi Asing, Jumlah Penduduk Dan Inflasi Terhadap Pengangguran Terdidik di Jawa Tengah Periode Tahun 1980-2011. *Economics Development Analysis Journal*.
- Wahba, G. (1990). *Spline Models for Observasional Data*. University of Winconsin at Madison, Phensylvania.
- Walpole, R. E. (1995). *Pengantar Statistika*. Jakarta: PT. Gramedia Pustaka Utama.
- Wei, W. W. (2006). *Time Series Univariate and Multivariate Methods*. Canada: Addison Wesley Publishing Company.Inc.
- Wijaya, A. (2019). Analisis Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Provinsi Aceh tahun 2015 dengan Regresi Nonparametrik Spline Truncated. *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya 2019 Terbitan I, 5 Oktober 2019*, 115-121.
- Wood, S. (2003). Thin Plate Regression Splines. *Journal of the Royal statistical Society: Series B (Statistical Methodology) Vol. 65, No. 1, pp , 95-114*.

LAMPIRAN

Lampiran 1: Data Penelitian**Data Tingkat Pengangguran Terbuka Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2020**

No	Kabupaten/Kota	Jumlah Perkotaan	Tingkat Pengangguran Terbuka
1	Kepulauan Selayar	292	2,44
2	Bulukumba	120	3,42
3	Bantaeng	73	4,27
4	Jeneponto	211	2,31
5	Takalar	162	4,16
6	Gowa	76	6,44
7	Sinjai	65	2,65
8	Maros	123	6,28
9	Pangkep	312	5,18
10	Barru	123	6,39
11	Bone	77	3,2
12	Soppeng	110	4,42
13	Wajo	53	4,33
14	Sidrap	73	5,91
15	Pinrang	126	4,19
16	Enrekang	101	2,44
17	Luwu	81	4,94
18	Tana Toraja	12	2,6

19	Luwu Utara	68	31
20	Luwu Timur	33	4,46
21	Toraja Utara	151	3,17
22	Makassar	22	15,92
23	Pare-pare	56	7,14
24	Palopo	63	10,37

Lampiran 2: Data Variabel Prediktor Tingkat Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2020

No	Kabupaten/Kota	x_1	x_2	x_3	x_4
1	Kepulauan Selayar	12,48	7,88	55,78	67,38
2	Bulukumba	7,1	7,67	51,98	68,99
3	Bantaeng	8,95	6,72	48,86	68,73
4	Jeneponto	14,58	6,59	51,86	64,26
5	Takalar	8,44	7,29	49,98	67,31
6	Gowa	7,38	8,19	50,97	70,14
7	Sinjai	9	7,75	58,1	67,6
8	Maros	9,74	7,73	51,25	69,86
9	Pangkep	13,96	7,66	51,58	68,72
10	Barru	8,26	8,23	54,23	71
11	Bone	10,68	7,15	52,94	66,06
12	Soppeng	7,59	7,81	51,13	68,67
13	Wajo	6,95	6,81	44,71	69,15
14	Sidrap	5,05	7,84	50,34	71,21
15	Pinrang	8,86	7,86	56,16	71,26
16	Enrekang	12,17	8,9	66,82	72,76
17	Luwu	12,65	8,24	60,84	70,51
18	Tana Toraja	12,1	8,26	65,98	68,75
19	Luwu Utara	13,41	7,79	58,19	69,57
20	Luwu Timur	6,85	8,8	54,82	73,22

21	Toraja Utara	12,01	7,96	71,38	68,33
22	Makassar	4,54	11,21	40,84	82,25
23	Pare-pare	5,44	10,45	48,69	77,86
24	Palopo	7,85	10,76	47,78	78,06

Keterangan:

x_1 : Persentase Penduduk Miskin

x_2 : Rata-rata Lama Sekolah

x_3 : *Dependency Ratio*

x_4 : Indeks Pembangunan Manusia