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	Dra. Nasrah Sirajang, M.Si.
NIP. 19620926 198702 2001	NIP. 19650519 199303 2002

Ketua Progam 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 kehadirat 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**, **Nuu**, **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

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan salah satu negara berkembang yang dalam pengelompokkan 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 fleksibelitas 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 fleksibelitas 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 siginifikan 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 menunjukan 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 popular 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 mareka yang baru menyelesaikan pendidikan berusaha mencari kerja sesuai dengan aspirasi mareka. Aspirasi mareka biasanya adalah bekerja di sektor modern atau di kantor, untuk mendapatkan pekerjaan itu mareka bersedia menunggu untuk beberapa lama, tidak tertutup kemungkinan mereka berusaha mencari perkerjaan 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.

$$Angkatan\ kerja = Jumlah\ orang\ yang\ bekerja\ +\ Jumlah\ yang\ tidak\ bekerja$$

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

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

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 $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}$ data yaitu sebagai berikut:

$$\bar{x} = \frac{\sum_{i=1}^{n} \boldsymbol{x_i}}{n} \tag{2.1}$$

dengan

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

$\boldsymbol{x_i}$: 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).

$$y_i = g(x_i) + \varepsilon_i, \qquad i = 1, 2, \dots, n \tag{2.2}$$

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

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.3}$$

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

$$\begin{aligned} y_i &= [1 \quad x_{1i} \quad x_{2i} \quad \cdots \quad x_{pi}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon_i \\ &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i, \qquad i = 1, 2, \dots, n \end{aligned} \tag{2.4}$$

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:

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

dimana

y_i : variabel respon

f : fungsi regresi nonparametrik

x_{pi} : variabel prediktor ke-p $(p = 1,2, \dots, j)$ pada pengamatan ke-i

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

2.4 *Likelihood Ratio Test* (LRT)

Metode *likelihod 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):

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

$$= \prod_{i=1}^{n} f(x_i , \theta)$$

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

$$= L(\theta | x) \quad (2.6)$$

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:

$$\ln L(x_1, x_2, \dots, x_n; \theta) \quad (2.7)$$

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

1. Menentukan fungsi likelihood
$$L(x_1, x_2, \dots, x_n; \theta) = (x_1; \theta)\, f(x_2; \theta)\ \dots\ f(x_n; \theta)$$
2. Membentuk logaritma natural likelihood
$$\ln L(x_1, x_2, \dots, x_n; \theta) = \ln(x_1; \theta)\, f(x_2; \theta)\ \dots\ f(x_n; \theta)$$
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),

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

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:

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

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

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

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

dengan:

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

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:

$$\boldsymbol{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{2.10}$$

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

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

Persamaan (2.10) dapat dituliskan dalam bentuk sebagai berikut:

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

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

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

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

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

$$f(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right)$$

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

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

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(\boldsymbol{\varepsilon_i})$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots K_r]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots K_r]\boldsymbol{\beta})\right) \quad (2.14)$$

Apabila persamaan (2.14) dibuat transformasi logaritma, akan didapatkan

$$l(\boldsymbol{\beta}) = logL(\boldsymbol{\beta})$$

$$= log\left((2\pi\sigma^2)^{-\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots K_r]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[K_1, K_2, \dots K_r]\boldsymbol{\beta})\right)\right)$$

$$= -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}) \quad (2.15)$$

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

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

$$= \frac{\partial\left(-\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}[K]'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}}$$

$$= -\frac{1}{2\sigma^2}(-2\boldsymbol{X}[K]'\boldsymbol{y} + 2\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta}) \quad (2.16)$$

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

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

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

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

Selanjutnya menentukan estimator model regresi nonparametrik spline *truncated*.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}[K]\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$= \boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-1}\boldsymbol{X}[K]'\boldsymbol{y} + \boldsymbol{\varepsilon}$$

$$= \boldsymbol{Ay} + \boldsymbol{\varepsilon} \tag{2.18}$$

dengan $\boldsymbol{A} = \boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-1}\boldsymbol{X}[K]'$ 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 biasnya 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 invarians terhadap transformasi.

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

$$GCV(K) = \frac{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\left[n^{-1}trace\left(\boldsymbol{I} - \boldsymbol{A}(K)\right)\right]^2} \tag{2.19}$$

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, **I** adalah matriks identitas, dan matriks $\boldsymbol{A} = \boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-1}\boldsymbol{X}[K]'$. 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 $N(0, \sigma^2\boldsymbol{I})$. Karena error $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\boldsymbol{I})$, maka $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ fungsi *likelihood* diberikan oleh:

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

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

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

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

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

$$= 0 - \frac{1}{2\sigma_\Omega^2}(-2\boldsymbol{X}[K]'\boldsymbol{y} + \boldsymbol{2X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta}_\Omega)$$

$$-\frac{1}{2\sigma_\Omega^2}\left(-2\boldsymbol{X}[K]'\boldsymbol{y} + 2\boldsymbol{X}[K]'\boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right) = 0$$

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

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

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

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

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

$$\frac{n}{2\hat{\sigma}_\Omega^2} = \frac{1}{2(\hat{\sigma}_\Omega^2)^2}\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)$$

$$\hat{\sigma}_\Omega^2 = \frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}{n} \tag{2.22}$$

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

$$\max_\Omega L\left(\boldsymbol{\beta}_\Omega, \sigma_\Omega^2\right) = L\left(\widehat{\Omega}\right) = L\left(\widehat{\boldsymbol{\beta}}_\Omega, \hat{\sigma}_\Omega^2\right)$$

$$L\left(\widehat{\Omega}\right) = \left(2\pi\hat{\sigma}_\Omega^2\right)^{-\frac{n}{2}}exp\left(-\frac{1}{2\hat{\sigma}_\Omega^2}\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)\right)$$

$$= \left(2\pi\hat{\sigma}_\Omega^2\right)^{-\frac{n}{2}}exp\left(-\frac{1}{2}\frac{n\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}\right)$$

$$= \left(2\pi\hat{\sigma}_\Omega^2\right)^{-\frac{n}{2}}exp\left(-\frac{n}{2}\right)$$

$$= \left(2\pi\hat{\sigma}_\Omega^2\right)^{-\frac{n}{2}}e^{-\frac{n}{2}} \tag{2.23}$$

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

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

$$= -\frac{n}{2}log(2\pi\sigma_\omega^2) - \frac{1}{2\sigma_\omega^2}(\boldsymbol{y}'\boldsymbol{y} - 2\sigma_\omega^2\boldsymbol{X}[K]'\boldsymbol{y} + \sigma_\omega^2\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta}_\omega) \tag{2.24}$$

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

$$F(\boldsymbol{\beta}_{\omega},\theta) = W(\boldsymbol{\beta}_{\omega}) + 2\theta\left(\boldsymbol{c}_j'\boldsymbol{\beta}_{\omega}\right) \tag{2.25}$$

dengan

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

Maka

$$F(\boldsymbol{\beta}_{\omega},\theta) = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}_{\omega}'\boldsymbol{X}[K]'\boldsymbol{y} + \boldsymbol{\beta}_{\omega}'\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta}_{\omega} + 2\theta\left(\boldsymbol{c}_j'\boldsymbol{\beta}_{\omega}\right) \tag{2.26}$$

Jika persamaan (2.26) diturunkan terhadap $\boldsymbol{\beta}_{\omega}$ maka akan menghasilkan persamaan:

$$\begin{aligned} \frac{\partial F(\boldsymbol{\beta}_{\omega},\theta)}{\partial \boldsymbol{\beta}_{\omega}} &= -2\boldsymbol{X}[K]'\boldsymbol{y} + 2\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta}_{\omega} + 2\theta\boldsymbol{c}_j \\ -2\boldsymbol{X}[K]'\boldsymbol{y} &+ 2\boldsymbol{X}[K]'\boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega} + 2\theta\boldsymbol{c}_j = 0 \\ 2\boldsymbol{X}[K]'\boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega} &= 2\boldsymbol{X}[K]'\boldsymbol{y} - 2\theta\boldsymbol{c}_j \\ \widehat{\boldsymbol{\beta}}_{\omega} &= (\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\left(\boldsymbol{X}[K]'\boldsymbol{y} - \theta\boldsymbol{c}_j\right) \\ &= (\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]'\boldsymbol{y} - (\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\theta\boldsymbol{c}_j \\ &= \widehat{\boldsymbol{\beta}}_{\Omega} - (\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\theta\boldsymbol{c}_j \end{aligned} \tag{2.27}$$

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

$$\frac{\partial F(\boldsymbol{\beta}_{\omega},\theta)}{\partial \theta} = 2\boldsymbol{c}_j'\boldsymbol{\beta}_{\omega}$$

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

$$\begin{aligned} \boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_{\omega} &= 0 \\ \boldsymbol{c}_j'\left(\widehat{\boldsymbol{\beta}}_{\Omega} - (\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\theta\boldsymbol{c}_j\right) &= 0 \\ \boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} - \boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\theta\boldsymbol{c}_j &= 0 \\ \boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\theta\boldsymbol{c}_j &= \boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \\ \theta &= \left(\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\right)^{-1}\boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \end{aligned} \tag{2.28}$$

Sehingga persamaan (2.27) akan menjadi

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

$$\widehat{\boldsymbol{\beta}}_{\omega} - \widehat{\boldsymbol{\beta}}_{\Omega} = (\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\left(\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\right)^{-1}\boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_{\Omega} \tag{2.29}$$

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

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

$$= -\frac{n}{2}\left(\frac{2\pi}{2\pi\sigma_{\omega}^2}\right) + \frac{1}{2(\sigma_{\omega}^2)^2}(\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}_{\omega})'(\boldsymbol{y} - \boldsymbol{X}[K]\boldsymbol{\beta}_{\omega})$$

$$-\frac{n}{2}\left(\frac{2\pi}{2\pi\hat{\sigma}_{\omega}^2}\right) + \frac{1}{2(\hat{\sigma}_{\omega}^2)^2}\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right) = 0$$

$$\frac{n}{2\hat{\sigma}_{\omega}^2} = \frac{1}{2(\hat{\sigma}_{\omega}^2)^2}\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)$$

$$\hat{\sigma}_{\omega}^2 = \frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)}{n} \tag{2.30}$$

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

$$\max_{\Omega} L(\boldsymbol{\beta}_{\omega}, \sigma_{\omega}^2) = L(\hat{\omega}) = L\left(\widehat{\boldsymbol{\beta}}_{\omega}, \hat{\sigma}_{\omega}^2\right)$$

$$L(\hat{\omega}) = (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}}exp\left(-\frac{1}{2\hat{\sigma}_{\omega}^2}\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)\right)$$

$$= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}}exp\left(-\frac{1}{2}\frac{n\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)}\right)$$

$$= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}}exp\left(-\frac{n}{2}\right)$$

$$= (2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}}e^{-\frac{n}{2}} \tag{2.31}$$

Selanjutnya diperoleh Rasio *Likelihood* sebagai berikut:

$$\lambda(x_1, \dots, x_h, y_i) = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{(2\pi\hat{\sigma}_{\omega}^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}}{(2\pi\hat{\sigma}_{\Omega}^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}} = \left(\frac{\hat{\sigma}_{\omega}^2}{\hat{\sigma}_{\Omega}^2}\right)^{-\frac{n}{2}}$$

$$\lambda(x_1, \dots, x_h, y_i) = \left(\frac{\frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\omega}\right)}{n}}{\frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_{\Omega}\right)}{n}}\right)^{-\frac{n}{2}}$$

$$= \left(\frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\omega\right)}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}\right)^{-\frac{n}{2}}$$

$$= \left(\frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\omega\right)}\right)^{\frac{2}{n}}$$

$$= \left(\frac{D_1}{D}\right)^{\frac{2}{n}} \tag{2.32}$$

dengan

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

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

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

Nilai $\widehat{\boldsymbol{\beta}}_\Omega$ akan disubstitusikan pada ruas kedua dan ketiga pada persamaan (2.33) yaitu $L = \left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\boldsymbol{X}[K]$ dan $\boldsymbol{M} = \boldsymbol{X}[K]'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)$ sebagai berikut:

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

Dan untuk M sebagai berikut:

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

Sehingga persamaan (2.33) akan menjadi:

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

Jika

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

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

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

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

maka akan diperoleh persamaan C sebagai berikut:

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

Sehingga persamaan D menjadi:

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

$$D = D_1 + D_2 \tag{2.35}$$

dengan

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

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

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

$$\lambda(x_1, \dots, x_h, y_i) = \left(\frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\omega\right)}\right)^{\frac{2}{n}} = \left(\frac{D_1}{D}\right)^{\frac{2}{n}}$$

$$= \left(\frac{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right) + \left(\boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\right)^{-1}\boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_\Omega}\right)^{\frac{2}{n}}$$

$$= \left(\frac{1}{1 + \dfrac{\left(\boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\right)^{-1}\boldsymbol{c}_j'\widehat{\boldsymbol{\beta}}_\Omega}{\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)}}\right)^{\frac{2}{n}}$$

$$= \left(\frac{1}{1 + \frac{D_2}{D_1}}\right)^{\frac{2}{n}} \tag{2.36}$$

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

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

dimana

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

dengan

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

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

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

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

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

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

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

$$\frac{\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}}{2\sigma^2} = \frac{1}{2\sigma^2}((\boldsymbol{X}[K]\boldsymbol{\beta})'\boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\left(\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\right)^{-1} \boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]'(\boldsymbol{X}[K]\boldsymbol{\beta}))$$

Akan dilakukan penjabaran pada pembilang sebagai berikut:

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

Sehingga dapat dinyatakan

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

Tahap kedua adalah dengan menjabarkan persamaan D_1 sebagai berikut:

$$D_1 = \left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)'\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)$$
$$= \left(\boldsymbol{y}' - \boldsymbol{X}[K]'\widehat{\boldsymbol{\beta}}'_\Omega\right)\left(\boldsymbol{y} - \boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega\right)$$
$$= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega - \widehat{\boldsymbol{\beta}}'_\Omega\boldsymbol{X}[K]'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'_\Omega\boldsymbol{X}[K]'\boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega$$
$$= \boldsymbol{y}'\boldsymbol{y} - 2\widehat{\boldsymbol{\beta}}'_\Omega\boldsymbol{X}[K]'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'_\Omega\boldsymbol{X}[K]'\boldsymbol{X}[K]\widehat{\boldsymbol{\beta}}_\Omega$$

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

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

dengan

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

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

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

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

$$tr(\boldsymbol{B}) = tr(\boldsymbol{I} - \boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]')$$
$$= tr(\boldsymbol{I}_n) - tr(\boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]')$$
$$= tr(\boldsymbol{I}_n) - tr\left(\boldsymbol{I}_{(q+r)+1}\right)$$
$$= n - \left((q+r)j + 1\right)$$
$$= n - (q+r)j - 1$$

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

$$\frac{\boldsymbol{\mu}'\boldsymbol{B}\boldsymbol{\mu}}{2\sigma^2} = \frac{1}{2\sigma^2}\Big((\boldsymbol{X}[K]\boldsymbol{\beta})'(\boldsymbol{I} - \boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]')(\boldsymbol{X}[K]\boldsymbol{\beta})\Big)$$

$$= \frac{1}{2\sigma^2}\Big(\boldsymbol{X}[K]'\boldsymbol{\beta}'\boldsymbol{X}[K]\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{X}[K]'\boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}(\boldsymbol{X}[K]\boldsymbol{\beta})\Big)$$

$$= \frac{1}{2\sigma^2}\big(\boldsymbol{\beta}'\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta} - (\boldsymbol{\beta}'\boldsymbol{X}[K]'\boldsymbol{X}[K]\boldsymbol{\beta})\big)$$

$$= 0$$

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

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

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

sehingga

$$\boldsymbol{AB} = \Big(\boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\big(\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{c}_j\big)^{-1}\boldsymbol{c}_j'(\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]'\Big) \times (\boldsymbol{I} - \boldsymbol{X}[K](\boldsymbol{X}[K]'\boldsymbol{X}[K])^{-\mathbf{1}}\boldsymbol{X}[K]')$$

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

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

$$= 0$$

Kajian pengujian hipotesis seperti yang diuraikan pada persamaan (2.20) sampai (2.41) telah dilakukan oleh Husni (2018). Selanjutnya, diperoleh bahwa $\boldsymbol{AB} = 0$ 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:

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

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

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

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:

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

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:

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

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 (*Sum Square*)**	**Kuadrat Tengah (*Mean Square*)**	**F**$_{\mathbf{hitung}}$
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 : minimal ada $\beta_l \neq 0; l = 1, 2, \ldots, q + r$

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

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

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 :

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

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 Kebaikan 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).

$$R^2 = \frac{SSR}{SST} \times 100\% \tag{2.46}$$

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.

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

Hipotesis untuk uji *Glejser*, yaitu:

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (residual identik)

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

Statistik uji yang digunakan :

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

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.

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

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).

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

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:

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

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 $\boldsymbol{A^2} = \boldsymbol{A}$.

2. **Teorema 1**

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

3. **Teorema 2**

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

4. **Teorema 3**

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

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

5. **Teorema 4**

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

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

6. **Teorema 5**

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

7. **Teorema 6**

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

8. **Teorema 7**

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

9. **Teorema 8**

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

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:

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