



TESIS - SS142501

**PENGUJIAN HIPOTESIS SIMULTAN PARAMETER
KOMPONEN PARAMETRIK DAN NONPARAMETRIK
DALAM MODEL REGRESI SEMIPARAMETRIK
CAMPURAN SPLINE *TRUNCATED* DAN *KERNEL***

(Studi Kasus: *Total Fertility Rate* kabupaten/kota di
Provinsi Jawa Timur)

ARIP RAMADAN
NRP. 062116 5001 0017

DOSEN PEMBIMBING
Dr. Dra. Ismaini Zain, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

**PROGRAM MAGISTER
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018**



TESIS - SS142501

**PENGUJIAN HIPOTESIS SIMULTAN PARAMETER
KOMPONEN PARAMETRIK DAN NONPARAMETRIK
DALAM MODEL REGRESI SEMIPARAMETRIK
CAMPURAN SPLINE *TRUNCATED* DAN *KERNEL*
(Studi Kasus: *Total Fertility Rate* kabupaten/kota di
Provinsi Jawa Timur)**

ARIP RAMADAN
NRP. 06211650010017

DOSEN PEMBIMBING
Dr. Dra. Ismaini Zain, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

PROGRAM MAGISTER
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018



THESIS - SS142501

***TESTING HYPOTHESES SIMULTAN
PARAMETERS PARAMETRIC AND
NONPARAMETRIC COMPONENTS IN THE MODEL
OF SEMIPARAMETRIC MIXED TRAVEL SPLINE
TRUNCATED AND KERNEL***

**(Case Study: Total Fertility Rate Region/City in East
Java)**

ARIP RAMADAN
NRP. 062116 5001 0017

SUPERVISORS
Dr. Dra. Ismaini Zain, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

MAGISTER PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018

**PENGUJIAN HIPOTESIS SIMULTAN PARAMETER KOMPONEN
PARAMETRIK DAN NONPARAMETRIK DALAM MODEL REGRESI
SEMPARAMETRIK CAMPURAN SPLINE TRUNCATED DAN KERNEL
(Studi Kasus: Total Fertility Rate kabupaten/kota di Provinsi Jawa Timur)**

**Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)**

**di
Institut Teknologi Sepuluh Nopember**

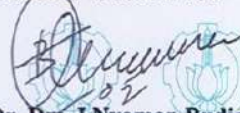
**oleh :
ARIP RAMADAN
NRP. 06211650010017**

**Tanggal Ujian : 18 Juli 2018
Periode Wisuda : September 2018**

Disetujui oleh:


1. **Dr. Dra. Ismaini Zain, M.Si**
NIP: 19600525 198803 2 001


(Pembimbing I)


2. **Prof. Dr. Drs. I Nvoman Budiantara, M.Si**
NIP: 19650603 198903 1 003

(Pembimbing II)


3. **Dr. Vita Ratnasari, S.Si, M.Si**
NIP: 19700910 199702 2 001

(Penguji)


4. **Dr. Puhadi, M.Sc**
NIP: 19620204 198701 1 001

(Penguji)

**Dekan
Fakultas Matematika, Komputasi dan Sains Data
Institut Teknologi Sepuluh Nopember**


Prof. Dr. Basuki Widodo, M.Sc
NIP: 19650605 198903 1 002

(Halaman ini sengaja dikosongkan)

**PENGUJIAN HIPOTESIS SIMULTAN PARAMETER
KOMPONEN PARAMETRIK DAN NONPARAMETRIK
DALAM MODEL REGRESI SEMIPARAMETRIK
CAMPURAN SPLINE *TRUNCATED* DAN *KERNEL*
(Studi Kasus: *Total Fertility Rate* kabupaten/kota di Provinsi Jawa
Timur)**

Nama Mahasiswa : Arip Ramadan
NRP : 062116 5001 0017
Pembimbing : Dr. Dra. Ismaini Zain, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

ABSTRAK

Persoalan pertumbuhan penduduk di Indonesia dari tahun ke tahun relatif sangat tinggi. Berdasarkan data Survei Demografi dan Kesehatan Indonesia (SDKI), terjadi peningkatan TFR yang sangat tinggi dan berdampak kepada tidak terkendalinya pertambahan jumlah penduduk di Indonesia. TFR di Provinsi Jawa Timur mengalami peningkatan yang sangat tinggi pada tahun 2012 mencapai 20%, oleh karena itu persoalan TFR di Provinsi Jawa Timur perlu dikaji. Perilaku pola TFR dikaitkan dengan variabel-variabel yang diduga berpengaruh yaitu *Unmet Need*, *Age Spesific Fertility Rates* (ASFR), Indeks Pembangunan Manusia (IPM) dan Angka Kematian Bayi (AKB) mempunyai pola yang sangat khusus. Pola hubungan antara TFR dengan *Unmet Need* cenderung linier. Sedangkan pola hubungan antara TFR dengan ASFR dan IPM, berbeda dengan *Unmet Need* yang cenderung nonlinier khususnya berubah-ubah seiring dengan meningkat atau menurunnya nilai dari variabel-variabel tersebut. Pada sisi lain pola hubungan antara TFR dengan AKB terlihat tidak mengikuti pola tertentu. Penelitian ini memodelkan pola hubungan antara TFR dengan variabel *Unmet Need* yang merupakan komponen parametrik, variabel ASFR dan IPM yang merupakan komponen spline *truncated* dan variabel AKB yang merupakan komponen *kernel*.

Untuk aplikasi uji hipotesis simultan komponen parameter model regresi semiparametrik spline *truncated* dan *kernel* diaplikasikan pada data *Total Fertility Rate*. Model terbaik yang diperoleh adalah menggunakan Spline *truncated* linier dengan kombinasi titik knot, berdasarkan pada nilai GCV optimum yang diperoleh yaitu sebesar 0,003964 dan koefisien determinasi sebesar 97,04%. Terdapat pengaruh yang signifikan dari variabel prediktor terhadap variabel respon setelah dilakukan pengujian hipotesis secara simultan dengan nilai F-hitung sebesar 101,85619.

Kata Kunci : Estimator Campuran Spline *Truncated* dan *Kernel*, *Total Fertility Rate*, Uji Hipotesis Simultan Parameter Parametrik dan Nonparametrik.

(Halaman ini sengaja dikosongkan)

**TESTING HYPOTHESES SIMULTAN PARAMETERS
PARAMETRIC AND NONPARAMETRIC COMPONENTS IN
THE MODEL OF SEMIPARAMETRIC MIXED TRAVEL SPLINE
TRUNCATED AND KERNEL**

(Case Study: Total Fertility Rate Region/City in East Java)

By : Arip Ramadan
Student Identity Number : 062116 5001 0017
Supervisor : Dr. Dra. Ismaini Zain, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

ABSTRACT

The problem of population growth in Indonesia year by year is relatively very high. Based on data from Indonesia Demographic and Health Survey there is a very high increase of TFR and impact to uncontrollable increase of population in Indonesia. TFR in East Java Province experienced a very high increase in 2012 reached 20%, therefore the issue of TFR in East Java Province needs to be studied. The behavior of TFR patterns is associated with the variables that are suspected to affect the Unmet Need, the Age Specific Fertility Rates (ASFR), the Human Development Index (HDI) and the Infant Mortality Rate (IMR) have a very special pattern. The relationship pattern between TFR and Unmet Need tends to be linear. While the pattern of relationship between TFR with ASFR and HDI, in contrast to the Unmet Need that tend to be nonlinear especially change as the increase or decrease value of these variables. On the other hand, the pattern of relationship between TFR and IMR appears not to follow a certain pattern. This study modeled the relationship between TFR and Unmet Need variable which is parametric component, ASFR and HDI variable which is spline truncated component and IMR variable which is kernel component.

For simultaneous hypothesis test application component parameter of spline truncated semiparametric regression model and kernel is applied to Total Fertility Rate data. The best model obtained is using linear spline linear with combination of knot point, based on GCV value of optimum obtained that is 0,003964 and coefficient of determination equal to 97,04%. There is a significant influence from the predictor variable to the response variable after hypothesis testing simultaneously with the F-count value of 101.85619.

Keywords : *Spline Truncated and Kernel Mixed Estimator, Total Fertility Rate Mapping, Parametric and Nonparametric Parameter Hypothesis Test,*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah memberikan rahmat, hidayah dan kekuatannya serta shalawat dan salam selalu tercurah pada Nabi Muhammad SAW atas suri tauladannya dalam kehidupan ini sehingga penulis dapat menyelesaikan Tesis yang berjudul **“Pengujian Hipotesis Simultan Parameter Komponen Parametrik dan Nonparametrik dalam model regresi semiparametrik campuran spline *truncated* dan *kernel* (Studi Kasus: Total Fertility Rate kabupaten/kota di Provinsi Jawa Timur)”**. terselesaikannya Tesis ini tidak lepas dari bantuan berbagai pihak, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Dr. Dra. Ismaini Zain, M.Si dan Prof. Dr. Drs. I Nyoman Budiantara, M.Si selaku dosen pembimbing yang telah banyak meluangkan waktunya untuk memberikan ilmu, bimbingan, dan arahnya kepada penulis dalam penyelesaian Tesis ini.
2. Dr. Vita Ratnasari, S.Si, M.Si dan Dr. Purhadi, M.Sc selaku dosen penguji yang telah memberi saran dan kritiknya demi kesempurnaan Tesis ini.
3. Dr. Suhartono, M.Sc selaku Ketua Departemen Statistika FMKSD-ITS.
4. Dr. rer pol. Heri Kuswanto, S.Si, M.Si selaku Ketua Program Magister Departemen Statistika FMKSD-ITS.
5. Prof. Drs. Nur Iriawan, MIKomp., Ph.D. selaku dosen wali yang telah memberikan semangat, motivasi serta masukan demi kelancaran dan terselesaikannya studi.
6. Segenap dosen pengajar dan para staff Departemen Statistika FMKSD-ITS yang memberikan bekal ilmu, memfasilitasi dan membantu penulis selama masa perkuliahan.
7. Orang tua yang senantiasa mendukung dan mendoakan penulis.
8. Teman-teman seperjuangan S2 2016 selama dua tahun ini, dalam Tesis dan semasa perkuliahan, terimakasih atas segala bantuan dan semangatnya.
9. Arina Bhuana Sari yang selalu mendukung dan memberikan motivasi.
10. Seluruh pihak yang telah banyak membantu penulis dan tidak dapat disebutkan satu per satu.

Dalam penulisan ini, penulis menyadari masih banyak kekurangan dalam penyusunan Tesis ini. Oleh karena itu, sangat diharapkan kritik dan saran yang membangun. Semoga Tesis ini bermanfaat bagi pembaca.

Surabaya, Juli 2018

Penulis

DAFTAR ISI

| | |
|---|------------------------------|
| LEMBAR PENGESAHAN | Error! Bookmark not defined. |
| ABSTRAK | iii |
| ABSTRACT | v |
| KATA PENGANTAR | vii |
| DAFTAR ISI | ix |
| DAFTAR GAMBAR | xii |
| DAFTAR LAMPIRAN | xiii |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 6 |
| 1.3 Tujuan Penelitian | 6 |
| 1.4 Manfaat Penelitian | 6 |
| 1.5 Batasan Masalah | 7 |
| BAB 2 TINJAUAN PUSTAKA | 9 |
| 2.1 Analisis Regresi | 9 |
| 2.2 Analisis Regresi Parametrik, Nonparametrik dan Semiparametrik | 10 |
| 2.3 Regresi Nonparametrik Spline <i>Truncated</i> | 13 |
| 2.4 Regresi Nonparametrik <i>Kernel</i> | 15 |
| 2.5 Regresi Semiparametrik Campuran Spline <i>Truncated</i> dan <i>Kernel</i> | 17 |
| 2.6 Pemilihan Titik Knot dan <i>bandwith</i> Optimal | 18 |
| 2.7 Pengujian Asumsi Residual | 19 |
| 2.7.1 Asumsi Independen | 19 |
| 2.7.2 Asumsi Identik | 19 |
| 2.7.3 Asumsi Distribusi Normal | 20 |
| 2.8 Koefisien Determinasi | 21 |
| 2.9 Angka Kelahiran Total (<i>Total Fertility Rate / TFR</i>) | 21 |
| BAB 3 METODOLOGI PENELITIAN | 23 |
| 3.1 Gambaran Objek Umum | 23 |
| 3.2 Sumber Data | 23 |
| 3.3 Variabel Penelitian | 23 |
| 3.4 Struktur Data | 24 |

| | | |
|--|---|-----------|
| 3.5 | Langkah-Langkah Penelitian | 25 |
| BAB 4 ANALISIS DAN PEMBAHASAN | | 27 |
| 4.1 | Model Regresi Semiparametrik Campuran Spline <i>Truncated</i> dan <i>Kernel</i> | 27 |
| 4.1.1 | Estimasi Kurva Regresi Semiparametrik Campuran Spline <i>Truncated</i> dan <i>Kernel</i> | 30 |
| 4.1.2 | Perumusan Uji Hipotesis | 35 |
| 4.1.3 | Pengujian Hipotesis..... | 39 |
| 4.1.4 | Mendapatkan Distribusi Statistik Uji | 46 |
| 4.2 | Aplikasi pada Data TFR di Provinsi Jawa Timur | 51 |
| 4.2.1 | Analisis Deskriptif..... | 51 |
| 4.2.2 | Pemodelan TFR Menggunakan Regresi Semiparametrik Campuran Spline <i>Truncated</i> dan <i>Kernel</i> | 53 |
| 4.2.2.1 | Penentuan Variabel Komponen Parametrik dan Komponen Nonparametrik | 53 |
| 4.2.3 | Model Regresi Semiparametrik Spline <i>Truncated</i> Linier | 55 |
| 4.2.3.1 | Memodelkan TFR di Provinsi Jawa Timur Menggunakan Semiparametrik Campuran Spline <i>Truncated</i> dan <i>Kernel</i> Satu Knot | 56 |
| 4.2.3.2 | Memodelkan TFR di Provinsi Jawa Timur dengan Komponen Spline <i>Truncated</i> Dua Titik Knot | 56 |
| 4.2.3.3 | Memodelkan TFR di Provinsi Jawa Timur dengan Komponen Spline <i>Truncated</i> Tiga Titik Knot..... | 57 |
| 4.2.3.4 | Memodelkan TFR di Provinsi Jawa Timur dengan Komponen Spline <i>Truncated</i> Kombinasi Titik Knot | 58 |
| 4.2.4 | Pemilihan Model Terbaik..... | 59 |
| 4.2.5 | Pengujian Asumsi Residual..... | 60 |
| 4.2.5.1 | Uji Identik | 60 |
| 4.2.5.2 | Uji Independen..... | 61 |
| 4.2.5.3 | Uji Distribusi Normal | 61 |
| 4.2.6 | Pengujian Signifikansi parameter secara Simultan | 62 |
| BAB 5 KESIMPULAN DAN SARAN | | 63 |
| 5.1 | Kesimpulan | 63 |
| 5.2 | Saran..... | 64 |
| DAFTAR PUSTAKA | | 65 |
| LAMPIRAN | | 67 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 3.1 Daftar Nama Daerah di Provinsi Jawa Timur | 23 |
| Tabel 3.2 Definisi Operasional Variabel | 24 |
| Tabel 3.3 Struktur Data Penelitian | 24 |
| Tabel 4.1 Statistik Deskriptif Variabel Respon dan Variabel Prediktor | 52 |
| Tabel 4.2 Komponen Parametrik dan Nonparametrik..... | 55 |
| Tabel 4.3 Perbandingan Nilai GCV dengan menggunakan Satu Titik Knot..... | 56 |
| Tabel 4.4 Perbandingan Nilai GCV dengan menggunakan Dua Titik Knot..... | 57 |
| Tabel 4.5 Perbandingan Nilai GCV dengan menggunakan Tiga Titik Knot | 58 |
| Tabel 4.6 Perbandingan Nilai GCV menggunakan Kombinasi Titik Knot..... | 59 |
| Tabel 4.7 Nilai GCV Minimum Pada Tiap Model | 59 |
| Tabel 4.8 ANOVA Uji Glejser | 60 |
| Tabel 4.9 ANOVA Hasil Regresi Semiparametrik campuran spline <i>truncated</i> dan <i>kernel</i> | 62 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 4.1 Plot antara TFR dengan Unmet Need | 53 |
| Gambar 4.2 <i>Scatter Plot</i> antara TFR dengan ASFR | 54 |
| Gambar 4.3 <i>Scatter plot</i> antara TFR dengan IPM | 54 |
| Gambar 4.4 <i>Scatter Plot</i> antara TFR dengan AKB..... | 55 |
| Gambar 4.5 Hasil Uji <i>Kolmogorov-Smirnov</i> | 62 |

DAFTAR LAMPIRAN

| | |
|--|----|
| Lampiran 1. Data yang Digunakan | 67 |
| Lampiran 2. Syntax R Model Semiparametrik Campuran Spline <i>kernel</i> | 68 |

(Halaman ini sengaja dikosongkan)

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan metode dipergunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon. Terdapat beberapa pendekatan yang ada dalam analisis regresi, yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik. Regresi parametrik digunakan jika bentuk kurva regresinya diketahui seperti linier, kuadratik, kubik, eksponensial dan lain sebagainya (Hardle, 1994). Jika pola data cenderung mengikuti model linear/kuadratik/kubik maka pendekatan regresi yang sesuai untuk data tersebut adalah regresi parametrik linear/kuadratik/kubik (Hardle, 1990). Apabila pola hubungan antara variabel prediktor dan variabel respon tidak diketahui bentuknya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data maka regresi nonparametrik merupakan pendekatan regresi yang dianjurkan untuk digunakan (Budiantara, 2009). Model regresi nonparametrik yang sering mendapat perhatian dari para peneliti adalah *Kernel* (Hardle, 1990), *Spline* (Wahba, 1990), Deret *Fourier* (Antoniadis, 1994) dan *Wavelets* (Antoniadis, 2007).

Regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik. Model regresi semiparametrik digunakan apabila variabel respon dapat memiliki pola hubungan tertentu dengan salah satu atau beberapa variabel prediktor, akan tetapi dengan variabel prediktor yang lain tidak diketahui bentuk pola hubungannya. Sering dijumpai kasus dimana masing-masing variabel prediktor mempunyai pola data yang berbeda. Apabila diestimasi hanya menggunakan satu bentuk estimator maka akan diperoleh estimator yang kurang sesuai dengan pola datanya. Akibatnya, estimasi model regresi yang dihasilkan tidak tepat dan cenderung mempunyai *error* yang besar. Penggunaan estimasi model kurva regresi yang sesuai dengan pola data akan lebih baik (Budiantara, Ratnasari, Ratna dan Zain, 2015). Oleh karena itu penggunaan estimator campuran dalam model regresi diperlukan untuk mengakomodir perbedaan pola data yang terjadi.

Pada beberapa penelitian, variabel respon dapat memiliki hubungan linier dengan sebagian variabel prediktor, dan memiliki pola hubungan yang tidak diketahui bentuk kurva regresinya dengan sebagian variabel prediktor yang lain. Pada kasus seperti ini, lebih disarankan untuk menggunakan pendekatan regresi semiparametrik (Wahba, 1990; Eubank, 1999; Budiantara, 2005).

Beberapa penelitian yang melibatkan estimator campuran dalam model regresi nonparametrik pernah dilakukan oleh (Sudiarsa dkk, 2015) dan Afifah (2017) yaitu melibatkan estimator *Spline Truncated* dan Deret Fourier. Sedangkan Budiantara (2015), Rory (2016), Purnomo (2016), Rismal (2016) dan Khusniawati (2017) melibatkan fungsi *Kernel* dan *Spline Truncated*. Sementara penelitian terkait regresi semiparametrik yang melibatkan estimator campuran belum banyak dilakukan, Hesikumalasari (2016) mengestimasi model semiparametrik campuran *Spline Truncated* dan *Kernel*, Fendy (2017) meneliti tentang estimasi interval parameter komponen parametrik dan spline dalam model regresi semiparametrik campuran spline dan *kernel*, sedangkan Nisa (2017) mengestimasi model semiparametrik campuran *Spline Truncated* dan Deret Fourier. Dari penelitian diatas hanya sebatas mencari estimasi titik dari kurva regresi dan belum dilakukan inferensi lebih lanjut terkait uji hipotesis simultan.

Diantara beberapa metode yang digunakan dalam regresi semiparametrik tersebut, dalam beberapa dekade terakhir model spline lebih banyak diminati dan mendapatkan perhatian dari para peneliti regresi nonparametrik, karena model spline merupakan model yang mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik (Eubank, 1999; Budiantara, 2009). Menurut Budiantara (2009), data yang memiliki pola berubah-ubah pada sub-sub interval tertentu sebaiknya dimodelkan dengan menggunakan metode spline. Spline merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Salah satu kelebihan spline adalah model ini cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam spline terdapat titik-titik knot, yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Eubank,1990; Budiantara,2009).

Spline sebagai pendekatan pola data dikenalkan oleh Whittaker pada tahun 1923. Spline dipopulerkan oleh Schoenberg pada tahun 1942. Sedangkan spline yang didasarkan pada suatu percobaan optimasi dikembangkan oleh Reinsch pada tahun 1967 (Wahba,1990). Pendekatan spline mempunyai suatu basis fungsi. Basis fungsi yang biasa digunakan antara lain spline *truncated* dan B-Spline. Spline *truncated* merupakan fungsi dimana terdapat perubahan pola perilaku kurva yang berbeda pada interval-interval yang berlainan. Beberapa peneliti yang telah mengkaji tentang fungsi spline dalam regresi nonparametrik diantaranya Crave dan Wahba (1979) melakukan kajian tentang metode *Cross Validation* (CV) untuk memilih parameter penghalus optimal dalam estimator spline. Selain itu, Wahba (1990) mengkaji tentang metode untuk memilih parameter penghalus optimal dalam estimator spline yaitu metode *Generalized Cross Validation* (GCV).

Inferensi statistik khususnya pengujian hipotesis terhadap parameter dalam model sangat penting dilakukan untuk mengetahui apakah variabel-variabel prediktor berpengaruh secara signifikan terhadap model. Wahba (1990) membuat suatu inferensi statistik khususnya interval konfidensi untuk model spline dalam regresi nonparameterik dengan pendekatan bayesian untuk smoothing spline, Zaki (2007) telah melakukan kajian tentang inferensi uji *Generalized Maximum Likelihood* (GML) untuk menguji hipotesis dalam model regresi nonparametrik spline dan menyelidiki perilaku uji GML yang dibandingkan dengan uji *Locally Most Powerful* (LMP) dengan menggunakan data simulasi. Tupen (2011) juga telah melakukan kajian tentang uji hipotesis dalam regresi nonparametrik spline *truncated*. Ruliana (2016) melakukan kajian tentang pengujian hipotesis simultan pada model spline pada *Structural Equation Modeling* (SEM) nonlinier. Sementara itu, pengembangan tentang pengujian hipotesis simultan khususnya pada regresi semiparametrik campuran spline *truncated* dan *kernel* belum pernah dilakukan. Kajian ini sangat penting karena dapat digunakan untuk mengetahui bentuk dari hipotesis, penurunan untuk mendapatkan statistik uji dan daerah penolakan dari pengujian hipotesis secara simultan khususnya pada regresi semiparametrik campuran spline *truncated* dan *kernel*.

Aplikasi pengujian hipotesis telah banyak dilakukan pada berbagai bidang keilmuan, salah satunya adalah pada bidang demografi. Penelitian bidang

demografi, diantaranya Angka Kelahiran Total (*Total Fertility Rate/TFR*) merupakan bidang yang sesuai untuk penerapan regresi semiparametrik campuran *spline truncated* dan *kernel*. Hal ini disebabkan karena hubungan antara variabel respon dengan sebagian variabel prediktor cenderung tidak diketahui pola hubungannya, namun dengan sebagian variabel prediktor lainnya membentuk hubungan linier.

Penduduk merupakan hal terpenting dalam menopang pembangunan suatu daerah karena merupakan subjek sekaligus objek dari pembangunan. Sebagai subjek pembangunan penduduk akan berperan dalam mempercepat tercapainya pembangunan ekonomi maupun sosial yang dapat berpengaruh terhadap meningkatnya kesejahteraan sosial, sedangkan sebagai objek dari pembangunan penduduk merupakan pihak yang mendapatkan hasil dari pembangunan suatu daerah.

Berdasarkan laporan yang dikeluarkan Survei Demografi dan Kesehatan Indonesia (SDKI) terakhir pada tahun 2012, terjadi peningkatan TFR nasional dari 2,41 pada tahun 2008 menjadi 2,6 pada tahun 2012. Berdasarkan laporan tersebut, hanya terdapat 10 provinsi yang mengalami penurunan tingkat fertilitasnya, sedangkan sisanya mengalami peningkatan. Peningkatan TFR yang dialami oleh provinsi lain berkisar antara 31 persen sampai 63 persen.

Provinsi Jawa Timur merupakan Provinsi yang mengalami penurunan TFR yang sangat signifikan sejak diberlakukannya kebijakan keluarga berencana (KB), yaitu TFR Provinsi Jawa Timur telah mencapai dibawah 2,1 pada tahun 2002. Namun TFR Provinsi Jawa Timur tahun 2012 mengalami peningkatan kembali yang signifikan sejak tahun 2002 yang berefek pada peningkatan kembali TFR Indonesia. TFR Jawa Timur mengalami peningkatan TFR lebih dari 20 persen yaitu dari 2,1 pada 2002 menjadi 2,6 pada tahun 2012.

Menurut SDKI (2012) *unmet need* adalah kebutuhan Keluarga Berencana yang tidak terpenuhi. Kebutuhan keluarga berencana yang belum terpenuhi (*unmet need*) didefinisikan sebagai kesenjangan antara niat wanita usia reproduksi dengan perilaku penggunaan kontrasepsi. Beberapa wanita yang menikah dan tidak menikah ingin menghindari kehamilan tetapi tidak menggunakan kontrasepsi. Definisi *unmet need* menurut *Demographic Health Survey* (DHS) adalah proporsi

wanita usia subur yang menikah atau hidup bersama (seksual aktif) yang tidak ingin punya anak lagi atau yang ingin menjarangkan kelahiran berikutnya dalam jangka waktu minimal 2 tahun tetapi tidak menggunakan alat atau cara kontrasepsi.

Unmet need merupakan pelayanan layanan kontrasepsi/keluarga berencana yang berkaitan dengan *demand* KB. *Demand* KB adalah niat atau motivasi individu atau pasangan untuk mengontrol TFR dimasa yang akan datang. Tingginya angka *Unmet Need* dalam permasalahan program KB mengindikasikan rendahnya *prevalensi* kontrasepsi yang berakibat tingginya angka kelahiran dan memicu terjadi ledakan penduduk yang berpengaruh terhadap tingginya TFR di Provinsi Jawa Timur. Pemerintah Provinsi Jawa Timur khususnya BKKBN memiliki target *Unmet Need* KB sebesar 7%. Namun Kenyataannya dari tahun 2011 – 2014 *Unmet Need* KB terus mengalami kenaikan hingga 10,48%.

Angka kelahiran menurut umur (*Age Spesific Fertility Rates/ASFR*) adalah angka yang menunjukkan banyaknya kelahiran per 1.000 perempuan pada kelompok umur tertentu. Menurut SDKI, angka rata-rata ASFR Provinsi Jawa Timur pada tahun 2007 adalah 66, sedangkan nasional adalah 51. Meskipun sudah berhasil diturunkan sampai angka 53 pada 2012, angkanya masih melampaui angka nasional yang mempunyai rata-rata 48. Rawannya angka ASFR Jatim bukan saja terdapat pada tingkat fertilitasnya yang tinggi dan waktunya yang panjang sampai di masa depan. Tapi, juga jumlah populasinya yang juga dominan.

Indeks pembangunan manusia (IPM) adalah pengukuran perbandingan dari harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua negara seluruh dunia. IPM dibentuk oleh 3 dimensi dasar yaitu: (1) umur panjang dan hidup sehat, (2) pengetahuan, dan (3) standar hidup layak. Mulai dari tahun 2013 IPM di Provinsi Jawa Timur menunjukkan peningkatan sampai tahun 2015 dengan peningkatan berturut-turut sebesar 67,55; 68,14 dan 68, 95.

Menurut data BPS, Angka Kematian Bayi di Jawa Timur terus menunjukkan penurunan yaitu tahun 2010 sebesar 29,99 kematian per 1000 kelahiran hidup, tahun 2011 menjadi 29,24 kematian bayi per 1000 kelahiran hidup, tahun 2012 menjadi 28,31 kematian bayi per 1000 kelahiran hidup, tahun 2013 menjadi 27,23 kematian per 1000 kelahiran hidup, tahun 2014 menjadi 26,55 kematian bayi per 1000 kelahiran hidup, dan tahun 2015 menjadi 25,82 kematian bayi per 1000

kelahiran hidup. Menurunnya tingkat kematian bayi tersebut disebabkan oleh semakin meningkatnya pengetahuan dan kesadaran penduduk dalam melaksanakan pola hidup sehat dan meningkatnya gizi keluarga serta diimbangi dengan peningkatan pelayanan kesehatan oleh pemerintah dengan jangkauan yang lebih luas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, pengujian hipotesis merupakan hal yang sangat penting dalam suatu penelitian. Selanjutnya pendekatan regresi semiparametrik campuran spline *truncated* dan *kernel* akan digunakan untuk memodelkan TFR kabupaten/kota di Provinsi Jawa Timur karena diduga memiliki bentuk kurva regresi antara variabel respon dengan sebagian tidak diketahui polanya dan sebagian cenderung mempunyai pola hubungan linier.

Rumusan permasalahan yang dapat diuraikan pada penelitian ini adalah sebagai berikut.

1. Bagaimana mendapatkan statistik uji untuk pengujian hipotesis secara simultan pada regresi semiparametrik campuran spline *truncated* dan *kernel*?
2. Bagaimana mendapatkan model regresi semiparametrik campuran spline *truncated* dan *kernel* pada data TFR kabupaten/kota di Provinsi Jawa Timur?

1.3 Tujuan Penelitian

Sesuai dengan permasalahan yang telah dipaparkan, Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mendapatkan statistik uji untuk pengujian hipotesis secara simultan pada model regresi semiparametrik campuran spline *truncated* dan *kernel*.
2. Mendapatkan model regresi semiparametrik campuran spline *truncated* dan *kernel* pada data TFR kabupaten/kota di Provinsi Jawa Timur.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan peneliti dari permasalahan mengenai TFR kabupaten/kota di Provinsi Jawa Timur dalam penelitian ini adalah:

1. Menambah wawasan keilmuan kepada peneliti khususnya dan masyarakat pada umumnya tentang regresi semiparametrik campuran spline *truncated* dan *kernel*.
2. Menghasilkan penurunan uji hipotesis regresi semiparametrik campuran spline *truncated* dan *kernel* sehingga dapat dijadikan tambahan referensi dalam melakukan analisis data.
3. Memberikan informasi secara lebih rinci mengenai TFR kabupaten/kota di Provinsi Jawa Timur sehingga dapat membantu pemerintah dalam menentukan kebijakan terkait pengendalian TFR.
4. Hasil penelitian diharapkan menjadi bahan masukan dan acuan untuk penelitian-penelitian selanjutnya.

1.5 Batasan Masalah

Batasan masalah yang digunakan agar penelitian dapat berjalan secara terarah serta dapat mencapai tujuan yang diinginkan adalah sebagai berikut.

1. Model regresi yang digunakan adalah semiparametrik campuran spline *truncated* dan *kernel*.
2. Metode yang digunakan untuk menentukan banyaknya titik knot dan lokasi titik knot optimum adalah metode *Generalized Cross Validation* (GCV).

(Halaman ini sengaja dikosongkan)

BAB 2

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan metode yang digunakan untuk menjelaskan hubungan antara variabel prediktor dengan variabel respon. Salah satu tujuan analisis regresi adalah untuk memperkirakan atau meramalkan nilai dari variabel respon apabila variabel prediktor ditetapkan nilainya (Drapper dan Smith,1992). Hubungan antara variabel respon dan prediktor untuk n pengamatan, dengan pengamatan $(x_i, y_i), i = 1, 2, \dots, n$ adalah sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i; i = 1, 2, \dots, n \quad (2.1)$$

dimana y_i merupakan y_i pada pengamatan ke- i , x_i adalah variabel prediktor x pada pengamatan ke- i , ε_i adalah *error* atau residual pada pengamatan ke- i yang merupakan variabel random independen dengan mean nol dan variansi konstan σ^2 , $f(x_i)$ adalah kurva regresi pada titik x_i (Eubank, 1999). Fungsi $f(x_i)$ dapat dituliskan dalam bentuk :

$$y_i = \tilde{x}_i^T \tilde{\beta} + \varepsilon_i \quad (2.2)$$

dimana $x_i^T = [1 \ x_{1i} \ x_{2i} \ \dots \ x_{pi}]$, $i = 1, 2, \dots, n$ sedangkan n adalah banyak data dan p adalah banyak variabel, sementara $\tilde{\beta} = [\beta_0 \ \beta_1 \ \dots \ \beta_p]^T$. Sehingga persamaannya menjadi:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i, i = 1, 2, \dots, n \quad (2.3)$$

Model regresi (2.2) dapat ditulis dalam bentuk matrik :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ 1 & x_{13} & x_{23} & \dots & x_{p3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.4)$$

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \tilde{\varepsilon} \quad (2.5)$$

dimana $\tilde{\mathbf{y}}$ merupakan vektor dari variabel respon berukuran $n \times 1$, \mathbf{X} merupakan matrik berukuran $(1 + p) \times n$ dan $\tilde{\boldsymbol{\beta}}$ adalah vektor parameter yang akan diestimasi berukuran $(1 + p) \times 1$, $\tilde{\boldsymbol{\varepsilon}}$ adalah vektor *error* random berukuran $n \times 1$. Estimator parameter model diperoleh berdasarkan berbagai metode yang telah dikenal dalam statistika, yaitu *Least Square* atau *Maximum Likelihood Estimator*. Parameter $\tilde{\boldsymbol{\beta}}$ diestimasi dengan metode *Least Square* yang meminimumkan $\tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}}$ sehingga diperoleh:

$$\hat{\tilde{\boldsymbol{\beta}}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \tilde{\mathbf{y}} \quad (2.6)$$

Dalam analisis regresi terdapat tiga pendekatan, yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik. Apabila pola hubungan antara variabel respon dan variabel prediktornya diketahui bentuknya maka digunakan pendekatan model regresi parametrik. Namun jika pola hubungan antara variabel respon dan variabel prediktornya tidak diketahui bentuknya maka digunakan pendekatan model regresi nonparametrik. Sedangkan apabila bentuk kurva regresi terdiri dari komponen parametrik dan komponen nonparametrik maka digunakan pendekatan regresi semiparametrik (Eubank, 1999; Budiantara, 2005).

2.2 Analisis Regresi Parametrik, Nonparametrik dan Semiparametrik

Pendugaan bentuk kurva regresi digunakan untuk menjelaskan hubungan antara variabel respon dengan variabel prediktor. Pendekatan pendugaan yang paling sering digunakan adalah pendekatan parametrik. Asumsi yang mendasari pendekatan regresi parametrik adalah bahwa bentuk kurva regresi dapat digambarkan oleh sekumpulan parameter tertentu atau dalam arti kata lain, dapat digambarkan dalam pola tertentu (Hardle, 1994). Hardle (1994) lebih lanjut menyatakan bahwa contoh tipe model parametrik adalah persamaan regresi polinomial yang parameter-parameternya merupakan koefisien dari variabel-variabel independen dari model regresi tersebut. Dalam model regresi parametrik, estimasi kurva regresi ekuivalen dengan melakukan estimasi terhadap parameter-parameter dalam model. Pendekatan model regresi parametrik memiliki sifat yang sangat baik dari sisi statistika inferensi seperti: sederhana, mudah interpretasinya, estimatornya tidak bias, efisien dan konsisten (Budiantara, 2009). Estimator

parameter-parameter model diperoleh berdasarkan berbagai metode yang telah dikenal dalam statistika yaitu, *Least Square* dan *Maximum Likelihood* (Wahba,1990; Lehman,1983).

Diberikan suatu model regresi parametrik linier multivariabel atau regresi linier berganda seperti persamaan 2.3 dengan y_i merupakan y pada pengamatan ke- i , $x_{i1}, x_{i2}, \dots, x_{ip}$ sebagai variabel prediktor, $\beta_0, \beta_1, \dots, \beta_p$ merupakan parameter yang tidak diketahui dan ε_i adalah *error* random yang independen berdistribusi normal dengan mean nol dan variansi σ^2 . Model regresi parametrik sama seperti pada persamaan 2.3.

Hardle (1994) menyatakan bahwa analisis regresi nonparametrik mulai dikenal sekitar abad ke- 19 tepatnya pada tahun 1857. Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan variabel respon yang tidak diketahui bentuk kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1999). Berdasarkan kenyataan tersebut, maka secara visual pola yang diberikan oleh variabel prediktor dan variabel respon tidak mempunyai pola yang mengikuti pola tertentu. Pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi, karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subyektivitas peneliti (Eubank, 1999).

Secara umum model regresi nonparametrik memiliki bentuk fungsi seperti bentuk regresi pada umum yaitu seperti persamaan 2.1 dengan y_i adalah variabel respon y pada pengamatan ke- i , sedangkan fungsi $f(x_i)$ merupakan kurva regresi pada titik x_i , dengan x_i sebagai variabel prediktor dan ε_i adalah *error* random yang diasumsikan berdistribusi normal independen dengan mean nol dan variansi σ^2 (Wahba, 1990). Regresi nonparametrik disebut regresi nonparametrik univariabel jika terdiri dari satu variabel respon dan satu variabel prediktor. Namun jika dalam regresi nonparametrik terdapat satu variabel respon dan lebih dari satu variabel prediktor, maka disebut regresi nonparametrik multivariabel (Budiantara, 2004). Jika diberikan data $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$ hubungan antara $(x_{i1}, x_{i2}, \dots, x_{ip})$ dan y_i dapat dituliskan sebagai berikut.

$$\begin{aligned}
y_i &= f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i \\
&= f(x_{1i}) + f(x_{2i}) + \dots + f(x_{pi}) + \varepsilon_i \\
&= \sum_{j=1}^p f(x_{ji}) + \varepsilon_i
\end{aligned} \tag{2.7}$$

dengan y_i adalah variabel respon, sedangkan fungsi f merupakan kurva regresi yang tidak diketahui bentuknya.

Dalam analisis regresi multiprediktor, variabel respon dapat memiliki pola hubungan tertentu dengan salah satu atau beberapa variabel prediktor, akan tetapi dengan variabel prediktor yang lain tidak diketahui bentuk hubungannya. Terhadap kejadian ini, Ruppert dkk. (2003) menyarankan untuk menggunakan pendekatan model regresi semiparametrik dikarenakan bahwa regresi semiparametrik merupakan gabungan antara regresi parametrik dengan regresi nonparametrik. Jika diberikan data $(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, y_i)$ maka hubungan antara $(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi})$ dan y_i dapat dituliskan sebagai berikut.

$$y_i = \sum_{j=1}^p f(x_{ji}) + \sum_{s=1}^q g(t_{si}) + \varepsilon_i \tag{2.8}$$

dengan y_i adalah variabel respon y pada pengamatan ke- i , sedangkan fungsi

$$\sum_{j=1}^p f(x_{ji}) \text{ merupakan komponen parametrik, dengan } \sum_{s=1}^q g(t_{si}) \text{ merupakan}$$

komponen nonparametrik dan ε_i adalah *error* random yang diasumsikan berdistribusi normal independen dengan mean nol dan variansi σ^2 . Jika fungsi komponen parametrik didekati dengan menggunakan fungsi linier maka persamaan 2.2 akan menjadi :

$$y_i = \tilde{x}_i^T \tilde{\beta} + \sum_{s=1}^q g(t_{si}) + \varepsilon_i \tag{2.9}$$

dimana y_i adalah variabel respon ke- i , $\tilde{x}_i^T = [1 \ x_{1i} \ x_{2i} \ \dots \ x_{pi}]$ adalah variabel prediktor untuk komponen parametrik, t_{si} adalah variabel prediktor untuk komponen nonparametrik, $\tilde{x}_i^T \tilde{\beta}$ merupakan komponen parametrik, $\tilde{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p] \in R^{p+1}$ parameter yang tidak diketahui, $g(t_{si})$ merupakan

fungsi komponen nonparametrik yang tidak diketahui dan ε_i adalah *error* dari model, dimana $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.

2.3 Regresi Nonparametrik Spline *Truncated*

Metode regresi nonparametrik memiliki fleksibilitas yang tinggi dan tidak memiliki asumsi mengenai bentuk pola hubungan antara variabel prediktor dan variabel respon (Eubank, 1999). Pendekatan regresi nonparametrik yang cukup populer adalah Spline. Spline merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Salah satu kelebihan Spline adalah model ini cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam Spline terdapat titik-titik knot, yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Eubank, 1999; Budiantara, 2009). Salah satu pendekatan yang dapat digunakan untuk memperoleh estimasi kurva regresi nonparametrik adalah pendekatan regresi spline *truncated*. Spline *truncated* merupakan pendekatan regresi nonparametrik yang banyak digunakan. Spline *truncated* merupakan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Salah satu kelebihan spline *truncated* adalah model ini cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam spline *truncated* terdapat titik-titik knot, yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data (Eubank, 1999; Budiantara, 2009).

Jika diberikan data berpasangan (t_i, y_i) , $i = 1, 2, \dots, n$ dimana y_i merupakan y pada pengamatan ke- i , sedangkan t_i adalah variabel prediktor yang mengikuti pola nonparametrik. Pola hubungan y_i dan t_i dapat dinyatakan dalam model regresi seperti persamaan persamaan dibawah ini:

$$y_i = \sum_{s=1}^q g(t_{si}) + \varepsilon_i \quad (2.10)$$

Selanjutnya jika bentuk kurva regresi $g(t_i)$ pada persamaan 2.10 dihipotesiskan dengan kurva regresi spline *truncated* dengan knot K_1, K_2, \dots, K_r maka:

$$g(t_i) = \sum_{j=0}^m \theta_j t_i^j + \sum_{k=1}^r \theta_{k+m} (t_i - K_k)_+^m \quad (2.11)$$

dengan θ adalah parameter-parameter model dan fungsi *truncated* $(t_i - K_k)_+^m$ diberikan oleh :

$$(t_i - K_k)_+^m = \begin{cases} (t_i - K_k)^m, & t_i \geq K_k \\ 0, & t_i < K_k \end{cases} \quad (2.12)$$

sehingga model regresi pada persamaan (2.12) menjadi:

$$y_i = \sum_{j=1}^m \theta_j t_i^j + \sum_{k=1}^r \theta_{k+m} (t_i - K_k)_+^m + \varepsilon_i \quad (2.13)$$

atau dapat ditulis dalam notasi matrik:

$$\tilde{y} = G(k)\tilde{\theta} + \tilde{\varepsilon} \quad (2.14)$$

dimana:

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

$$G(k) = \begin{bmatrix} t_1^1 & \cdots & t_1^m & (t_1 - K_1)_+^m & (t_1 - K_2)_+^m & \cdots & (t_1 - K_r)_+^m \\ t_2^1 & \cdots & t_2^m & (t_2 - K_1)_+^m & (t_2 - K_2)_+^m & \cdots & (t_2 - K_r)_+^m \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_n^1 & \cdots & t_n^m & (t_n - K_1)_+^m & (t_n - K_2)_+^m & \cdots & (t_n - K_r)_+^m \end{bmatrix}, \quad (2.15)$$

$$\tilde{\theta} = [\theta_0 \quad \theta_1 \quad \cdots \quad \theta_m \quad \theta_{m+1} \quad \theta_{m+2} \quad \cdots \quad \theta_{m+r}]^T,$$

$$\tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Estimator untuk parameter model bisa diperoleh berdasarkan berbagai metode, salah satunya dengan menggunakan *Least Square*. Dengan menggunakan metode *Least Square* didapatkan estimator sebagai berikut.

$$\hat{\theta} = [G(k)^T G(k)]^{-1} G(k)^T \tilde{y} \quad (2.16)$$

dengan demikian estimasi kurva regresi $\hat{g}(t_i)$ diperoleh:

$$\begin{aligned} \hat{g}(t_i) &= G(k)\hat{\theta} \\ &= G(k)[G(k)^T G(k)]^{-1} G(k)^T \tilde{y} \\ &= A(k)\tilde{y} \end{aligned} \quad (2.17)$$

dimana:

$$A(t) = G(k)[G(k)^T G(k)]^{-1} G(k)^T \quad (2.18)$$

2.4 Regresi Nonparametrik *Kernel*

Estimator *kernel* diperkenalkan pertama kali oleh Rosenblatt (1956) dan Parzen (1962) sehingga namanya disebut estimator densitas *kernel* Rosenblatt-Parzen (Hardle, 1990). Estimator *kernel* merupakan perkembangan dari estimator histogram. Estimator *kernel* merupakan estimator linier yang serupa dengan estimator lainnya, perbedaannya adalah metode *kernel* lebih khusus dalam penggunaan *bandwidth* (Eubank, 1988). Beberapa kelebihan estimator *kernel* adalah fleksibel, bentuk matematisnya mudah, dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara dan Mulianah, 2007).

Diberikan pasangan pengamatan independen $(z_i, y_i), i = 1, 2, \dots, n$ dimana y_i merupakan y pada pengamatan ke- i , sedangkan z_i adalah variabel prediktor. Hubungan antara y_i dan z_i dapat dimodelkan dalam bentuk berikut.

$$y_i = \sum_{k=1}^r h(z_{ki}) + \varepsilon_i \quad (2.19)$$

dimana kurva regresi $h(z_i)$ merupakan kurva yang tidak diketahui bentuknya.

Menurut Hardle (1994), kurva $\sum_{k=1}^r h(z_{ki})$ pada model nonparametrik dapat diestimasi dengan pendekatan *kernel* yang didasarkan pada fungsi densitas *kernel*.

Kurva $\sum_{k=1}^r h(z_{ki})$ pada model 2.19 dapat diestimasi menggunakan estimator *kernel*

Nadaraya-Watson.

$$\hat{h}_\alpha(z) = n^{-1} \sum_{i=1}^n \left[\frac{K_\alpha(z - z_i)}{\sum_{j=1}^n K_\alpha(z - z_j)} \right] y_i = n^{-1} \sum_{i=1}^n W_{\alpha i}(z) y_i \quad (2.20)$$

dimana $\hat{h}_\alpha(z)$ adalah fungsi taksiran regresi *kernel*, α merupakan lebar *bandwidth*, $W_{\alpha i}(z)$ merupakan fungsi pembobot.

$$W_{\alpha i}(z) = \frac{K_\alpha(z - z_i)}{n^{-1} \sum_{j=1}^n K_\alpha(z - z_j)} \quad (2.21)$$

dan $K_\alpha(z - z_i)$ merupakan fungsi *kernel*

$$K_\alpha(z - z_i) = \frac{1}{\alpha} K\left(\frac{z-z_i}{\alpha}\right) \quad (2.22)$$

Estimator *kernel* sangat bergantung pada fungsi *kernel* K dan *bandwidth* α . Dalam bentuk matrik dapat dituliskan:

$$\begin{aligned} \begin{bmatrix} \hat{h}_\alpha(z_1) \\ \hat{h}_\alpha(z_2) \\ \vdots \\ \hat{h}_\alpha(z_n) \end{bmatrix} &= \begin{bmatrix} n^{-1} \sum_{i=1}^n W_{\alpha i}(z_1) y_i \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(z_2) y_i \\ \vdots \\ n^{-1} \sum_{i=1}^n W_{\alpha i}(z_n) y_i \end{bmatrix} \\ &= \begin{bmatrix} n^{-1} W_{\alpha 1}(z_1) y_1 + \cdots + n^{-1} W_{\alpha n}(z_1) y_n \\ n^{-1} W_{\alpha 1}(z_2) y_1 + \cdots + n^{-1} W_{\alpha n}(z_2) y_n \\ \vdots \\ n^{-1} W_{\alpha 1}(z_n) y_1 + \cdots + n^{-1} W_{\alpha n}(z_n) y_n \end{bmatrix} \quad (2.23) \\ &= \begin{bmatrix} n^{-1} W_{\alpha 1}(z_1) & n^{-1} W_{\alpha 2}(z_1) & \cdots & n^{-1} W_{\alpha n}(z_1) \\ n^{-1} W_{\alpha 1}(z_2) & n^{-1} W_{\alpha 2}(z_2) & \cdots & n^{-1} W_{\alpha n}(z_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1} W_{\alpha 1}(z_n) & n^{-1} W_{\alpha 2}(z_n) & \cdots & n^{-1} W_{\alpha n}(z_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \end{aligned}$$

sehingga persamaan (2.24) dapat dituliskan dalam notasi matrik berikut ini.

$$\hat{\tilde{h}}_\alpha(z) = D(\alpha) \tilde{y} \quad (2.24)$$

dimana vektor $\hat{\tilde{h}}(z)$ berukuran $n \times 1$, vektor \tilde{y} berukuran $n \times 1$ dan matrik $D(\alpha)$ berukuran $n \times n$.

Fungsi *kernel* K adalah fungsi yang bernilai riil, kontinu, terbatas dan simetris, dan integralnya sama dengan satu atau $\int K(v) dv = 1$. Menurut Hardle (1994), fungsi *kernel* dapat berupa:

1. *Kernel* Gaussian : $\left(K(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right), -\infty < v < \infty \right)$
2. *Kernel* Uniform : $\left(K(v) = \frac{1}{2}, |v| \leq 1 \right)$
3. *Kernel* Epanechnikov : $\left(K(v) = \frac{3}{4}(1 - v^2), |v| \leq 1 \right)$
4. *Kernel* Kuadrat : $\left(K(v) = \frac{15}{16}(1 - v^2)^2, |v| \leq 1 \right)$

dimana $v_i = \left(\frac{z-z_i}{\alpha}\right)$, $i = 1, 2, \dots, n$. z adalah variabel prediktor, z_i adalah nilai ke- i variabel prediktor dan α adalah *bandwidth*. Estimator *kernel* sangat bergantung pada pemilihan *bandwidth* α yang berfungsi untuk mengontrol *smoothness* dari kurva estimasi. Pemilihan *bandwidth* yang tepat merupakan hal yang sangat penting dalam regresi *kernel*. Beberapa kelebihan estimator *kernel* yaitu fleksibel, bentuk matematisnya mudah dan dapat mencapai tingkat kekonvergenan yang relatif cepat (Budiantara dan Mulianah, 2007).

2.5 Regresi Semiparametrik Campuran Spline *Truncated* dan *Kernel*

Regresi semiparametrik merupakan gabungan antara komponen parametrik dan komponen nonparametrik (Budiantara, 2009). Dalam beberapa kasus, dapat ditemukan hubungan antara variabel respon dengan salah satu variabel prediktor bersifat linier, tetapi hubungan dengan variabel prediktor yang lain tidak diketahui pola hubungannya. Variabel yang memiliki pola data yang diketahui atau terdapat informasi sebelumnya mengenai pola datanya digolongkan pada komponen parametrik. Sedangkan variabel yang tidak diketahui pola datanya digolongkan pada komponen nonparametrik.

Diberikan data berpasangan $(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, z_{1i}, z_{2i}, \dots, z_{ri}, y_i)$ maka model semiparametrik dirumuskan sebagai berikut.

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, z_{1i}, z_{2i}, \dots, z_{ri}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.25)$$

dengan kurva diasumsikan aditif, maka diperoleh:

$$y_i = \sum_{j=1}^p f(x_{ji}) + \sum_{s=1}^q g(t_{si}) + \sum_{k=1}^r h(z_{ki}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.26)$$

dimana y_i adalah variabel respon, $\sum_{j=1}^p f(x_{ji})$ adalah komponen parametrik,

$\sum_{s=1}^q g(t_{si})$ adalah komponen nonparametrik spline, $\sum_{k=1}^r h(z_{ki})$ adalah komponen nonparametrik *kernel*, dan ε_i adalah *error* random yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian σ^2 . Kurva regresi

komponen parametrik $\sum_{j=1}^p f(x_{ji})$ pada persamaan (2.26) dihampiri dengan fungsi

linier 2.1 dan dapat dituliskan dalam bentuk matrik 2.4. Selanjutnya kurva regresi

$\sum_{s=1}^q g(t_{si})$ dihipotesiskan dihipotesiskan dengan fungsi spline 2.10 dan dituliskan dalam

bentuk matrik 2.15. Kurva regresi $\sum_{k=1}^r h(z_{ki})$ dihipotesiskan dengan fungsi *kernel* 2.19

dan dapat dituliskan dalam bentuk matrik 2.23.

Model regresi semiparametrik pada persamaan 2.26 dapat disajikan dalam bentuk matrik sebagai berikut.

$$\tilde{y} = \mathbf{X}\tilde{\beta} + G(k)\tilde{\theta} + D(\alpha)\tilde{y} + \tilde{\varepsilon} \quad (2.27)$$

dimana \tilde{y} merupakan vektor dari variabel respon berukuran $n \times 1$, \mathbf{X} merupakan matrik komponen parametrik berukuran $n \times (p + 1)$, vektor $\tilde{\beta}$ berukuran $(1 + p) \times 1$, G merupakan matrik komponen spline berukuran $n \times (m + r)$, vektor $\tilde{\theta}$ berukuran $(m + r) \times 1$, matrik $D(\alpha)$ berukuran $n \times n$ dan $\tilde{\varepsilon}$ adalah vektor *error* random berukuran $n \times 1$.

2.6 Pemilihan Titik Knot dan *bandwith* Optimal

Dalam regresi nonparametrik maupun semiparametrik dengan pendekatan Spline, hal penting yang berperan dalam mendapatkan estimator Spline adalah pemilihan titik knot yang optimal. Sementara itu pada *kernel* bergantung pada *bandwidth*. *Bandwidth* α adalah parameter pemulus (*smoothing*) yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. Titik knot maupun *bandwidth* yang terlalu kecil akan menghasilkan kurva yang sangat kasar dan fluktuatif, sebaliknya titik knot maupun *bandwidth* yang terlalu besar atau lebar akan menghasilkan kurva yang sangat mulus, tetapi tidak sesuai dengan pola data (Hardle,1994). Pemilihan titik knot k dan *bandwidth* yang optimum menggunakan GCV di definisikan sebagai berikut.

$$GCV(k, \alpha) = \frac{MSE(k, \alpha)}{(n^{-1}trace[(I - M(k, \alpha))]^2)} \quad (2.28)$$

dimana $M(k, \alpha)$ merupakan persamaan untuk menaksir estimator semiparametrik campuran spline *truncated* linear dan *kernel* dan $MSE(k, \alpha)$ merupakan *Mean*

Square Error pada model nonparametrik spline *truncated* dan *kernel* yang didapatkan dari persamaan berikut.

$$MSE(k, \alpha) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.29)$$

2.7 Pengujian Asumsi Residual

Pada model regresi semiparametrik Spline *truncated* linier diasumsikan bahwa *error* random berdistribusi normal independen dengan mean nol dan variansi σ^2 (Wahba, 1990). Oleh karena itu sebelum melakukan analisis dan mengambil keputusan dari hasil pemodelan maka dilakukan uji asumsi residual terlebih dahulu. Uji asumsi residual yang dilakukan adalah uji independen, uji identik dan uji normalitas.

2.7.1 Asumsi Independen

Pemeriksaan asumsi independensi residual digunakan untuk mendeteksi korelasi antara residual. Salah satu metode yang dapat digunakan adalah dengan menggunakan statistik uji *Durbin-Watson*. Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \rho = 0$ (tidak terdapat *autokorelasi*)

$H_1 : \rho \neq 0$ (Terdapat *autokorelasi*)

dengan statistik uji: $d_{hitung} = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$

Adapun kesimpulannya adalah sebagai berikut :

- i. Tolak H_0 jika $d_{hitung} < d_{L, \frac{\alpha}{2}}$ dimana d_L adalah batas nilai bawah tabel.
- ii. Gagal tolak H_0 jika $d_{hitung} > d_{U, \frac{\alpha}{2}}$ dimana d_U adalah batas nilai atas tabel.
- iii. Jika $d_{L, \frac{\alpha}{2}} < d_{hitung} < d_{U, \frac{\alpha}{2}}$ maka tidak ada keputusan.

2.7.2 Asumsi Identik

Uji identik dapat juga disebut uji homogenitas varians residual. Homogenitas varians residual didasarkan pada sifat $E(\varepsilon_i) = 0$ dimana $V(\varepsilon_i) = \sigma^2$. Pada kondisi ini, varians residual konstan. Pelanggaran terhadap asumsi ini disebut heteroskedastis yaitu keadaan dimana variansi residual tidak homogen. Hal ini menyebabkan estimasi koefisien kurang akurat/tidak efisien (Gujarati, 1992). Cara

yang dapat digunakan dalam mendeteksi adanya heteroskedastisitas diantaranya adalah dengan menggunakan Uji Glejser dengan hipotesis sebagai berikut.

$$H_0 : \sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_n = \sigma^2$$

$$H_1 : \text{Minimal ada satu } \sigma^2_i \neq \sigma^2, i = 1, 2, \dots, n$$

Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hitung} = \frac{MSR}{MSE} = \left(\frac{\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}_i|)^2}{p-1} \right) / \left(\frac{\sum_{i=1}^n (|e_i| - |\hat{e}_i|)^2}{n-p} \right) \quad (2.31)$$

Daerah penolakan H_0 jika nilai $F_{hitung} > F_{\alpha;(p,n-p)}$ atau $P_{value} < \alpha$. Apabila pada kesimpulan dihasilkan penolakan H_0 , maka dapat dinyatakan bahwa artinya terdapat minimal satu $\sigma^2_i \neq \sigma^2$ dan itu berarti terdapat heteroskedastisitas.

2.7.3 Asumsi Distribusi Normal

Pengujian hipotesis melibatkan distribusi tertentu, yaitu F dan t -Student, kedua distribusi tersebut digunakan untuk menguji signifikansi parameter baik secara simultan maupun parsial. Oleh karena itu diperlukan pengujian distribusi normal pada residualnya, sehingga jika residual tidak memenuhi asumsi normal maka pengujian parameter menjadi tidak akurat. Cara mendeteksi apakah residual berdistribusi normal dapat dilihat pada normal *probability plot residual*. Apabila plot yang dihasilkan membentuk garis lurus maka residual dari model regresi tersebut cenderung mengikuti distribusi normal. Pengujian distribusi normal dapat dilakukan dengan metode Uji *Kolmogorov-Smirnov* yang juga dikenal dengan uji kesesuaian model (*Goodness of fit test*). Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : F_0(x) = F(x) \text{ (Residual berdistribusi normal)}$$

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

Statistik uji :

$$D = maks |F_0(x) - S_N(x)| \quad (2.32)$$

$F_0(x)$ adalah fungsi distribusi frekuensi kumulatif teoritis sedangkan $S_N(x) = k/n$ merupakan fungsi peluang kumulatif yang diobservasi dari satu sampel random dengan N observasi. k adalah banyaknya observasi yang sama atau kurang dari x .

Kesimpulan untuk menolak H_0 jika $|D| > q_{(1-\alpha)}$ dengan nilai $q_{(1-\alpha)}$ berdasarkan tabel *Kolmogorov-Smirnov*.

2.8 Koefisien Determinasi

Koefisien Determinasi (R^2) merupakan ukuran ketelitian atau ketepatan model regresi, atau besarnya kontribusi prediktor terhadap respon. Rumus Koefisien Determinasi sebagai berikut.

$$R^2 = \frac{SSR}{SST} = \frac{(\hat{Y} - \bar{Y})(\hat{Y} - \bar{Y})}{(Y - \bar{Y})(Y - \bar{Y})} \quad (2.33)$$

Besaran nilai R^2 tidak akan negatif dan batasannya adalah $0 \leq R^2 \leq 1$ (Gujarati dan Porter, 2015).

2.9 Angka Kelahiran Total (*Total Fertility Rate* / TFR)

TFR yaitu rata-rata jumlah anak yang dilahirkan oleh seorang wanita sejak awal usia subur sampai akhir masa reproduksinya (BKKBN, 2017). Namun ada catatan yang harus diperhatikan yaitu:

- a. Tidak ada seorang perempuan yang meninggal sebelum mengakhiri masa reproduksinya.
- b. Tingkat *fertilitas* menurut umur tidak berubah pada periode waktu tertentu (Mantra, 2006).

Menurut Mantra (2006), TFR menggambarkan riwayat fertilitas dari sejumlah perempuan hipotesis selama masa reproduksinya. Hal ini sesuai dengan riwayat kematian dari tabel kematian penampang lintang (*cross sectional life table*). Dalam praktek TFR dikerjakan dengan menjumlahkan tingkat *fertilitas* perempuan menurut umur, apabila umur tersebut berjenjang lima tahunan, dengan asumsi bahwa *fertilitas* menurut umur tunggal sama dengan rata-rata tingkat *fertilitas* kelompok umur lima tahunan.

(Halaman ini sengaja dikosongkan)

BAB 3

METODOLOGI PENELITIAN

3.1 Gambaran Objek Umum

Penelitian ini menggunakan data kabupaten/kota di Provinsi Jawa Timur. Wilayah Provinsi Jawa Timur terdiri atas 38 daerah meliputi 29 kabupaten dan 9 kota. Secara rinci dapat dinyatakan dengan Tabel 3.1 berikut ini.

Tabel 3.1 Daftar Nama Daerah di Provinsi Jawa Timur

| NO | NAMA KAB/KOTA | NO | NAMA KAB/KOTA | NO | NAMA KAB/KOTA |
|-----------|--------------------------|-----------|--------------------------|-----------|--------------------------|
| 1 | Pacitan | 14 | Pasuruan | 27 | Sampang |
| 2 | Ponorogo | 15 | Sidoarjo | 28 | Pemekasan |
| 3 | Trenggalek | 16 | Mojokerto | 29 | Sumenep |
| 4 | Tulungagung | 17 | Jombang | 30 | Kota Kediri |
| 5 | Blitar | 18 | Nganjuk | 31 | Kota Blitar |
| 6 | Kediri | 19 | Madiun | 32 | Kota Malang |
| 7 | Malang | 20 | Magetan | 33 | Probolinggo |
| 8 | Lumajang | 21 | Ngawi | 34 | Kota Pasuruan |
| 9 | Jember | 22 | Bojonegoro | 35 | Kota Mojokerto |
| 10 | Banyuwangi | 23 | Tuban | 36 | Kota Madiun |
| 11 | Bondowoso | 24 | Lamongan | 37 | Kota Surabaya |
| 12 | Situbondo | 25 | Gresik | 38 | Kota Batu |
| 13 | Probolinggo | 26 | Bangkalan | | |

3.2 Sumber Data

Penelitian ini menggunakan data sekunder tahun 2015 dari publikasi Badan Kependudukan dan Keluarga Berencana Nasional Provinsi Jawa Timur dengan unit observasi meliputi 38 kabupaten/kota dan Badan Pusat Statistik Provinsi Jawa Timur.

3.3 Variabel Penelitian

Variabel respon yang digunakan dalam penelitian ini berupa data kategorik, yaitu TFR menurut kabupaten/kota di Provinsi Jawa Timur tahun 2015 dan variabel-variabel yang diduga mempengaruhi TFR. Variabel-variabel tersebut dapat diuraikan dalam Tabel 3.2.

Tabel 3.2 Definisi Operasional Variabel

| Var | Nama Variabel | Definisi Operasional |
|-------|---|--|
| Y | TFR (<i>Total Fertility Rate</i>) | Rata-rata jumlah anak yang dilahirkan oleh seorang wanita sejak awal usia subur sampai akhir masa reproduksinya |
| X_1 | <i>Unmet Need</i> | angka persentase yang menunjukkan kebutuhan keluarga berencana yang tidak terpenuhi atau proporsi wanita usia subur yang menikah atau hidup bersama (seksual aktif) yang tidak ingin mempunyai anak lagi atau yang ingin menja rangkan kelahiran berikutnya dalam jangka waktu minimal 2 tahun tetapi tidak menggunakan alat atau cara kontrasepsi pada suatu kabupaten/kota |
| X_2 | ASFR (<i>Age Spesific Fertility Rate</i>) | Banyaknya kelahiran per 1000 wanita pada kelompok usia tertentu antara 15-49 tahun |
| X_3 | Indeks Pembangunan Manusia | Pengukuran harapan hidup, melek huruf, pendidikan dan standar hidup untuk semua negara seluruh dunia. |
| X_4 | Angka Kematian Bayi | Banyaknya kematian bayi pada satu tahun tertentu per 1000 kelahiran hidup pada tahun yang sama |

3.4 Struktur Data

Unit observasi yang digunakan dalam penelitian ini adalah sebanyak 38 kabupaten/kota di Provinsi

Jawa Tengah dengan empat buah variabel prediktor. Sehingga, struktur data penelitian yang digunakan adalah sebagai berikut.

Tabel 3.3 Struktur Data Penelitian

| Obs | Respon | Prediktor | | | | | | | | | | | |
|----------|----------|---------------------|------------|----------|------------|------------|------------------------|----------|------------|------------|------------|----------|------------|
| | | Komponen Parametrik | | | | | Komponen Nonparametrik | | | | | | |
| | y | x_1 | x_2 | \dots | x_p | t_1 | t_2 | \dots | t_q | z_1 | z_2 | \dots | z_r |
| 1 | y_1 | $x_{1;1}$ | $x_{2;1}$ | \dots | $x_{p;1}$ | $t_{1;1}$ | $t_{2;1}$ | \dots | $t_{q;1}$ | $z_{1;1}$ | $z_{2;1}$ | \dots | $z_{r;1}$ |
| 2 | y_2 | $x_{1;2}$ | $x_{2;2}$ | \dots | $x_{p;2}$ | $t_{1;2}$ | $t_{2;2}$ | \dots | $t_{q;2}$ | $z_{1;2}$ | $z_{2;2}$ | \dots | $z_{r;2}$ |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| 38 | y_{38} | $x_{1;38}$ | $x_{2;38}$ | \dots | $x_{p;38}$ | $t_{1;38}$ | $t_{2;38}$ | \dots | $t_{q;38}$ | $z_{1;38}$ | $z_{2;38}$ | \dots | $z_{r;38}$ |

3.5 Langkah-Langkah Penelitian

Tujuan pertama dalam penelitian ini adalah melakukan kajian tentang pengujian hipotesis simultan pada model regresi semiparametrik campuran spline *truncated* dan *kernel*. Tahapan penelitian yang dilakukan untuk menyelesaikan tujuan pertama ini adalah sebagai berikut.

1. Diberikan respon y_i dengan variabel prediktor komponen parametrik $x_{1i}, x_{2i}, \dots, x_{pi}$ sedangkan variabel prediktor $t_{1i}, t_{2i}, \dots, t_{qi}$ dan $z_{1i}, z_{2i}, \dots, z_{ri}$ merupakan komponen nonparametrik. Hubungan antara respon dan prediktor diasumsikan mengikuti model regresi semiparametrik yang bersifat additif.

$$y_i = \mu(x_{1i}, x_{2i}, \dots, x_{pi}, t_{1i}, t_{2i}, \dots, t_{qi}, z_{1i}, z_{2i}, \dots, z_{ri}) + \varepsilon_i$$

$$= \sum_{j=1}^p f_j(x_{ji}) + \sum_{s=1}^q g_s(t_{si}) + \sum_{k=1}^r h_k(z_{ki}) + \varepsilon_i, i = 1, 2, \dots, n, \varepsilon_i \sim N(0, \sigma^2)$$

2. Menghampiri kurva-kurva regresi tersebut dengan fungsi-fungsi sebagai berikut.
 - a. Menghampiri $f_j(x_{ji})$ dengan menggunakan fungsi linier pada persamaan 2.1
 - b. Menghampiri fungsi $g_s(t_{si})$ menggunakan fungsi Spline *Truncated* pada persamaan 2.12
 - c. Menghampiri fungsi $h_k(z_{ki})$ menggunakan fungsi *Kernel* pada persamaan 2.21.
3. Tuliskan model regresi semiparametrik campuran spline *truncated* dan *kernel* dalam bentuk matrik.
 - a. Komponen Parametrik: $\tilde{f}(x) = \mathbf{X}\tilde{\beta}$,
 - b. Komponen Nonparametrik Spline *Truncated*: $\tilde{g}(t) = \mathbf{G}(k)\tilde{\theta}$,
 - c. Komponen Nonparametrik *Kernel*: $\tilde{h}(z) = \mathbf{D}(\alpha)\tilde{y}$.
4. Melakukan estimasi untuk kurva regresi \tilde{f} , \tilde{g} dan \tilde{h} .
5. Merumuskan uji hipotesis simultan untuk parameter komponen parametrik dalam model regresi semiparametrik campuran spline *truncated* dan *kernel* sebagai berikut.

$$H_0 : \mathbf{C}^T \tilde{\gamma} = \boldsymbol{\tau}$$

$$H_1 : \mathbf{C}^T \tilde{\gamma} \neq \boldsymbol{\tau}$$

6. Menentukan ruang parameter dibawah $H_0(\omega)$

$$\omega = \left\{ \gamma = (\beta_0, \dots, \beta_p, \theta_{11}, \dots, \theta_{(r+m)1}, \dots, \theta_{1q}, \dots, \theta_{(m+r)q}), \sigma_\omega^2 \mid \mathbf{C}^T \tilde{\gamma} = \boldsymbol{\tau} \right\}$$

7. Mencari fungsi likelihood dibawah ruang $H_0(\omega)$

8. Memaksimumkan fungsi likelihood dibawah $\omega(L(\omega))$
9. Menentukan ruang parameter dibawah $H_0(\Omega)$
10. $\Omega = \{\gamma = (\beta_0, \dots, \beta_p, \theta_{11}, \dots, \theta_{(r+m)1}, \dots, \theta_{1q}, \dots, \theta_{(m+r)q}), \sigma_\Omega^2\}$
11. Mencari fungsi likelihood dibawah ruang Ω
12. Memaksimumkan fungsi likelihood dibawah $\Omega(L(\Omega))$
13. Membuat *ratio likelihood*

$$\lambda(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}, \tilde{y}) = \frac{L(\omega)}{L(\Omega)}, 0 \leq \lambda \leq 1$$

14. Mencari statistik uji berdasarkan langkah (13).
15. Mendapatkan distribusi statistik uji.
16. Menentukan daerah penolakan hipotesis H_0 , melalui $\lambda(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}, y) < k$, untuk suatu konstanta k .

Tujuan kedua dari penelitian ini adalah memodelkan TFR dengan menggunakan regresi semiparametrik spline *truncated* dan *kernel* selanjutnya melakukan pengujian hipotesis terhadap parameter-parameter model. Langkah-langkah yang dilakukan adalah:

1. Membuat plot variabel respon dengan semua variabel-variabel prediktor.
2. Menentukan variabel komponen parametrik dan nonparametrik spline *truncated* dan *kernel*.
3. Memodelkan variabel respon dan variabel prediktor menggunakan estimator campuran spline *truncated* dan *kernel* dalam regresi semiparametrik.
4. Memodelkan data dengan regresi semiparametrik campuran spline *truncated* dan *kernel* dengan satu, dua, tiga, dan kombinasi knot.
5. Memilih titik knot, parameter dan bandwidth α yang optimal berdasarkan metode GCV.
6. Melakukan pengujian signifikansi parameter secara simultan dan parsial.
7. Melakukan pengujian terhadap asumsi independen, identik dan distribusi normal untuk residual.

BAB 4

ANALISIS DAN PEMBAHASAN

Berdasarkan tujuan penelitian maka pada bagian ini membahas tentang model dan estimasi regresi semiparametrik campuran spline *truncated* dan *kernel*. Selanjutnya dilakukan uji hipotesis secara simultan untuk mengetahui hubungan antara variabel prediktor dengan variabel respon yang di terapkan pada data TFR di Provinsi Jawa Timur tahun 2015.

4.1 Model Regresi Semiparametrik Camuran Spline *Truncated* dan *Kernel*

Diberikan data berpasangan $(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}, z_i, \dots, z_{ri}, y_i)$ dan hubungan data berpasangan tersebut mengikuti mengikuti model regresi semiparametrik campuran spline *truncated* dan *kernel* sebagai berikut.

$$y_i = \mu(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}, z_i, \dots, z_{ri}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (4.1)$$

dengan kurva regresi μ pada persamaan (4.1) diasumsikan bersifat aditif, sehingga dapat dituliskan menjadi :

$$\begin{aligned} y_i &= f_1(x_{1i}) + \dots + f_p(x_{pi}) + g_1(t_{1i}) + \dots + g_q(t_{qi}) + h_1(z_{1i}) + \dots + h_r(z_{ri}) + \varepsilon_i \\ &= \sum_{j=1}^p f_j(x_{ji}) + \sum_{s=1}^q g_s(t_{si}) + \sum_{k=1}^r h_k(z_{ki}) + \varepsilon_i \end{aligned} \quad (4.2)$$

dimana y_i adalah variabel respon, $\sum_{j=1}^p f_j(x_{ji})$ adalah komponen parametrik,

$\sum_{s=1}^q g_s(t_{si})$ adalah komponen nonparametrik spline *truncated*, $\sum_{k=1}^r h_k(z_{ki})$ adalah

komponen non-parametrik *kernel*, dan ε_i adalah *error* random yang diasumsikan identik, inde-penden dan berdistribusi normal.

Kurva regresi $\sum_{j=1}^p f_j(x_{ji})$ merupakan komponen parametrik, dihampiri

dengan fungsi linier dan dapat ditulis sebagai berikut.

$$\begin{aligned} \sum_{j=1}^p f_j(x_{ji}) &= \beta_0 + \beta_j x_{ji} + \dots + \beta_p x_{pi} \\ &= \sum_{j=0}^p \beta_j x_{ji} \end{aligned} \quad (4.3)$$

Dari persamaan fungsi linier 4.3 diatas berlaku untuk $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$, dapat dituliskan dalam bentuk matrik sebagai berikut.

$$\begin{bmatrix} f_j(x_{j1}) \\ f_j(x_{j2}) \\ \vdots \\ f_j(x_{jn}) \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \dots + \beta_p x_{p1} \\ \beta_0 + \beta_1 x_{12} + \dots + \beta_p x_{p2} \\ \vdots \\ \beta_0 + \beta_1 x_{1n} + \dots + \beta_p x_{pn} \end{bmatrix}$$

Selanjutnya dalam notasi matriks dapat ditulis menjadi :

$$\tilde{f}(x) = \mathbf{X}\tilde{\beta} \quad (4.4)$$

dimana :

$$\tilde{f}(x) = \begin{bmatrix} f_j(x_{j1}) \\ f_j(x_{j2}) \\ \vdots \\ f_j(x_{jn}) \end{bmatrix}_{n \times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{12} & \dots & x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{pn} \end{bmatrix}_{(p+1) \times n} \quad \text{dan} \quad \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1) \times 1}$$

Kurva regresi $g_s(t_{si})$ merupakan komponen nonparametrik, dihipotesis dengan fungsi spline *truncated* derajat bebas m dan titik-titik knot K_1, K_2, \dots, K_r sebagai berikut.

$$\sum_{s=1}^q g_s(t_{si}) = \sum_{s=1}^q \left(\sum_{j=1}^m \theta_{sj} t_{si} + \sum_{k=1}^r \theta_{s(m+k)} (t_{si} - K_{sk})_+^m \right) \quad (4.5)$$

Dari persamaan 4.5 diatas dapat dituliskan fungsi spline linier dengan $m = 1$ untuk $s = 1$ sebagai berikut.

$$\begin{aligned} g_1(t_i) &= \theta_{11} t_i + \sum_{k=1}^r \theta_{1(1+k)} (t_i - K_{1k})_+ \\ &= \theta_{11} t_i + \theta_{12} (t_i - K_{11})_+ + \dots + \theta_{1(1+r)} (t_i - K_{1r})_+ \end{aligned} \quad (4.6)$$

dengan $i = 1, 2, \dots, n$ maka menjadi :

$$\begin{bmatrix} g_1(t_{11}) \\ g_1(t_{12}) \\ \vdots \\ g_1(t_{1n}) \end{bmatrix} = \begin{bmatrix} \theta_{11} t_{11} + \theta_{12} (t_{11} - K_{s1})_+ + \dots + \theta_{1(1+r)} (t_{11} - K_{1r})_+ \\ \theta_{11} t_{12} + \theta_{12} (t_{12} - K_{s1})_+ + \dots + \theta_{1(1+r)} (t_{12} - K_{1r})_+ \\ \vdots \\ \theta_{11} t_{1n} + \theta_{12} (t_{1n} - K_{s1})_+ + \dots + \theta_{1(1+r)} (t_{1n} - K_{1r})_+ \end{bmatrix}$$

atau dapat dituliskan dalam notasi matriks

$$\tilde{g}_1(t_1) = \mathbf{G}(k_1)\tilde{\theta}_1$$

dimana :

$$\tilde{g}_1(t_1) = \begin{bmatrix} g_1(t_{11}) \\ g_1(t_{12}) \\ \vdots \\ g_1(t_{1n}) \end{bmatrix}, \mathbf{G}(k_1) = \begin{bmatrix} t_{11} & (t_{11} - K_{11})_+ & \cdots & (t_{11} - K_{1r})_+ \\ t_{12} & (t_{12} - K_{11})_+ & \cdots & (t_{12} - K_{1r})_+ \\ \vdots & \vdots & \ddots & \vdots \\ t_{1n} & (t_{1n} - K_{11})_+ & \cdots & (t_{1n} - K_{1r})_+ \end{bmatrix}, \tilde{\theta}_1 = \begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1(1+r)} \end{bmatrix}$$

hal ini berlaku juga untuk $s = 1, 2, \dots, q$ sehingga diperoleh :

$$\begin{aligned} \sum_{s=1}^q g_s(t_{si}) &= g_1(t_{1i}) + g_2(t_{2i}) + \dots + g_q(t_{qi}) \\ &= \mathbf{G}(k_1)\tilde{\theta}_1 + \mathbf{G}(k_2)\tilde{\theta}_2 + \dots + \mathbf{G}(k_q)\tilde{\theta}_q \\ &= \begin{bmatrix} \mathbf{G}(k_1) & \mathbf{G}(k_2) & \dots & \mathbf{G}(k_q) \end{bmatrix} \begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \vdots \\ \tilde{\theta}_q \end{bmatrix} \\ &= \mathbf{G}(\tilde{k})\tilde{\theta} \end{aligned} \quad (4.7)$$

Sementara itu, kurva regresi $h_k(z_{ki})$ merupakan komponen nonparametrik, dihipotesis dengan fungsi *kernel* dan dapat ditulis sebagai berikut.

$$\sum_{k=1}^r h_k(z_{ki}) = \sum_{k=1}^r \left(n^{-1} \sum_{i=1}^n W_{ki}(Z_{ki}) y_i \right) \quad (4.8)$$

Dari persamaan 4.8 diatas dapat dituliskan fungsi kernel untuk $k = 1$ sebagai berikut.

$$\hat{h}_1(z_{1i}) = n^{-1} \sum_{i=1}^n W_{1i}(Z_{1i}) y_i$$

dimana :

$$W_{1i}(z_i) = \frac{\frac{1}{\alpha} K\left(\frac{z_i - z_j}{\alpha}\right)}{n^{-1} \sum_{j=1}^n \frac{1}{\alpha} K\left(\frac{z_i - z_j}{\alpha}\right)}$$

dengan $i=1,2,\dots,n$ maka menjadi :

$$\begin{bmatrix} \hat{h}_1(z_{11}) \\ \hat{h}_1(z_{12}) \\ \vdots \\ \hat{h}_1(z_{1n}) \end{bmatrix} = \begin{bmatrix} n^{-1}W_{\alpha_1}(Z_{11})y_1 + n^{-1}W_{\alpha_2}(Z_{11})y_2 + \dots + n^{-1}W_{\alpha_n}(Z_{11})y_n \\ n^{-1}W_{\alpha_1}(Z_{12})y_1 + n^{-1}W_{\alpha_2}(Z_{12})y_2 + \dots + n^{-1}W_{\alpha_n}(Z_{12})y_n \\ \vdots \\ n^{-1}W_{\alpha_1}(Z_{1n})y_1 + n^{-1}W_{\alpha_2}(Z_{1n})y_2 + \dots + n^{-1}W_{\alpha_n}(Z_{1n})y_n \end{bmatrix}$$

atau dapat dituliskan dalam notasi matriks

$$\hat{h}_1(z_i) = \mathbf{D}(\alpha_1) \tilde{y}$$

dimana $\hat{h}_\alpha(z_i)$ merupakan estimator kurva regresi kernel Nadaraya-Watson, parameter α merupakan parameter *bandwidth*, fungsi kernel K yang digunakan adalah kernel Gaussian. Dari persamaan fungsi kernel di atas, untuk $i=1$ sampai n , didapat persamaan : :

$$\hat{h}_1(z_i) = \begin{bmatrix} \hat{h}_1(z_{11}) \\ \hat{h}_1(z_{12}) \\ \vdots \\ \hat{h}_1(z_{1n}) \end{bmatrix}, \tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{D}(\alpha_1) = \begin{bmatrix} n^{-1}W_{\alpha_1}(Z_{11}) & n^{-1}W_{\alpha_2}(Z_{11}) & \dots & n^{-1}W_{\alpha_n}(Z_{11}) \\ n^{-1}W_{\alpha_1}(Z_{12}) & n^{-1}W_{\alpha_2}(Z_{12}) & \dots & n^{-1}W_{\alpha_n}(Z_{12}) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_{\alpha_1}(Z_{1n}) & n^{-1}W_{\alpha_2}(Z_{1n}) & \dots & n^{-1}W_{\alpha_n}(Z_{1n}) \end{bmatrix}$$

hal ini berlaku juga untuk $k=1,2,\dots,r$ sehingga diperoleh :

$$\begin{aligned} \sum_{k=1}^r h_k(z_{ki}) &= \hat{h}_1(z_{1i}) + \hat{h}_1(z_{2i}) + \dots + \hat{h}_1(z_{ri}) \\ &= \mathbf{D}(\alpha_1) \tilde{y} + \mathbf{D}(\alpha_2) \tilde{y} + \dots + \mathbf{D}(\alpha_r) \tilde{y} \\ &= \mathbf{D}(\tilde{\alpha}) \tilde{y} \end{aligned} \tag{4.9}$$

4.1.1 Estimasi Kurva Regresi Semiparametrik Campuran Spline *Truncated* dan *Kernel*

Estimator semiparametrik campuran spline *truncated* dan *kernel* dapat diperoleh dengan menggunakan metode *ordinary least square* (OLS). Dari model pada persamaan 4.11 diperoleh :

$$\begin{aligned} \varepsilon &= \tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} - \mathbf{D}(\tilde{\alpha})\tilde{y} \\ &= \tilde{y} - \mathbf{D}(\tilde{\alpha})\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \\ &= (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \end{aligned} \tag{4.10}$$

selanjutnya didapat:

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \tilde{\varepsilon}^T \tilde{\varepsilon} \\ &= \left[(\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \mathbf{X} \tilde{\beta} - \mathbf{G}(\tilde{k}) \tilde{\theta} \right]^T \left[(\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \mathbf{X} \tilde{\beta} - \mathbf{G}(\tilde{k}) \tilde{\theta} \right] \end{aligned} \quad (4.11)$$

kemudian diperoleh:

$$\sum_{i=1}^n \varepsilon_i^2 = [\tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T - \tilde{\beta}^T \mathbf{X}^T - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T] [(\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \mathbf{X} \tilde{\beta} - \mathbf{G}(\tilde{k}) \tilde{\theta}] \quad (4.12)$$

jika disederhanakan menjadi:

$$\begin{aligned} \tilde{\varepsilon}^T \tilde{\varepsilon} &= \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T \mathbf{X} \tilde{\beta} - \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T \mathbf{G}(\tilde{k}) \tilde{\theta} - \\ &\quad \tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{X} \tilde{\beta} + \tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ &\quad \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{X} \tilde{\beta} + \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta} \end{aligned} \quad (4.13)$$

selanjutnya diperoleh:

$$\begin{aligned} \tilde{\varepsilon}^T \tilde{\varepsilon} &= \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \\ &\quad \tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{X} \tilde{\beta} + \tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} - \tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ &\quad \tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} + \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta} \end{aligned} \quad (4.14)$$

dengan sedikit penjabaran diperoleh:

$$\begin{aligned} \tilde{\varepsilon}^T \tilde{\varepsilon} &= \tilde{y}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - 2\tilde{\beta}^T \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \\ &\quad 2\tilde{\theta}^T \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \tilde{\beta}^T \mathbf{X}^T \mathbf{X} \tilde{\beta} + 2\tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} + \\ &\quad \tilde{\theta}^T \mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta} \\ &= Q(\tilde{\beta}, \tilde{\theta}). \end{aligned} \quad (4.15)$$

Selanjutnya Derivatif Parsial terhadap $\tilde{\beta}$ didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\beta}} = 0 - 2\mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + 2\mathbf{X}^T \mathbf{X} \tilde{\beta} + 2\tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}. \quad (4.16)$$

Sementara itu Derivatif parsial terhadap $\tilde{\theta}$ didapat :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0 - 2\mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + 2\mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta} + 2\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \tilde{\theta} \quad (4.17)$$

didapat persamaan:

$$-2\mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + 2\mathbf{X}^T \mathbf{X} \tilde{\beta} + 2\tilde{\beta}^T \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta} = 0$$

selanjutnya dapat ditulis:

$$\mathbf{X}^T \mathbf{X} \tilde{\beta} = \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}$$

sehingga diperoleh:

$$\begin{aligned}\beta &= (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \tilde{\theta}\end{aligned}\quad (4.18)$$

selanjutnya untuk:

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0$$

didapat persamaan:

$$-2\mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + 2\mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta} + 2\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \theta = 0$$

selanjutnya dapat ditulis:

$$\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}) \theta = \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta}$$

sehingga diperoleh:

$$\begin{aligned}\theta &= (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} (\mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta}) \\ &= (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta}\end{aligned}\quad (4.19)$$

kemudian mensubstitusikan $\tilde{\theta}$ kedalam persamaan 4.20, didapat:

$$\begin{aligned}\hat{\beta} &= \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right] \left[(\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + \right. \\ &\quad \left. - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta} \right] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} + \\ &\quad (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta}\end{aligned}$$

dengan menyederhanakan didapat:

$$\begin{aligned}\hat{\beta} &- (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \tilde{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}} \\ &- (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{\mathbf{y}}\end{aligned}$$

selanjutnya dapat ditulis menjadi:

$$\begin{aligned}\left[\mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k}) \tilde{\theta} \mathbf{X} \right] \hat{\beta} &= \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ &\left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{\mathbf{y}}\end{aligned}$$

kemudian diperoleh:

$$\hat{\beta} = \left[\mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y}$$

estimator $\hat{\beta}$ dapat ditulis menjadi:

$$\hat{\beta} = \mathbf{A}(\tilde{k}, \tilde{\alpha}) \tilde{y}, \quad (4.20)$$

dimana :

$$\mathbf{A}(\tilde{k}, \tilde{\alpha}) = \left[\mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right]$$

selanjutnya, substitusikan $\tilde{\beta}$ ke dalam persamaan 4.24 didapat:

$$\hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \right. \\ \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \hat{\theta} \right)$$

kemudian diperoleh:

$$\hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \hat{\theta}$$

dengan sedikit penjabaran didapat:

$$\hat{\theta} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y}$$

persamaan diatas dapat diuraikan menjadi:

$$\left(\mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right) \hat{\theta} = (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y}$$

selanjutnya diperoleh:

$$\hat{\theta} = \left[\mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y}$$

estimator $\hat{\theta}$ dapat ditulis menjadi:

$$\hat{\theta} = \mathbf{B}(\tilde{k}, \tilde{\alpha}) \tilde{y}, \quad (4.21)$$

dimana:

$$\mathbf{B}(\tilde{k}, \tilde{\alpha}) = \left[\mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \\ - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right]$$

berdasarkan $\hat{\beta}$ didapat estimator kurva parametrik linear:

$$\begin{aligned} \hat{f}(x) &= \mathbf{X} \hat{\beta} \\ &= \mathbf{X} \left[\mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ &\quad \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y} \\ &= \mathbf{X} \mathbf{A}(\tilde{k}, \tilde{\alpha}) \tilde{y} \end{aligned} \quad (4.22)$$

dimana:

$$\mathbf{X} \mathbf{A}(\tilde{k}, \tilde{\alpha}) = \mathbf{X} \left[\mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \right. \\ \left. - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right]$$

berdasarkan $\hat{\theta}$ didapat estimator kurva regresi *Spline truncated* linear:

$$\begin{aligned} \tilde{g}(t) &= \mathbf{G}(\tilde{k}) \hat{\theta} \\ &= \mathbf{G}(\tilde{k}) \left[\mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \\ &\quad - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y} \\ &= \mathbf{G} \mathbf{B}(\tilde{k}, \tilde{\alpha}) \tilde{y} \end{aligned} \quad (2.23)$$

dimana

$$\mathbf{G} \mathbf{B}(\tilde{k}, \tilde{\alpha}) = \mathbf{G}(\tilde{k}) \left[\mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) + \\ - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right].$$

estimator campuran *Spline truncated* linear dan kernel dalam regresi semiparametrik diperoleh:

$$\begin{aligned}
\hat{\mu}(x, t, z) &= \hat{f}(x) + \hat{g}(t) + \hat{h}(z) \\
&= \mathbf{XA}(\tilde{k}, \tilde{\alpha})\tilde{y} + \mathbf{GB}(\tilde{k}, \tilde{\alpha})\tilde{y} + \mathbf{D}(\tilde{\alpha})\tilde{y} \\
&= (\mathbf{XA}(\tilde{k}, \tilde{\alpha}) + \mathbf{GB}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha}))\tilde{y} \\
&= ((\mathbf{XA} + \mathbf{GB})(\tilde{k}) + \mathbf{D})\tilde{\alpha})\tilde{y} \\
&= \mathbf{M}(\mathbf{X}, \mathbf{G}, \mathbf{D}, \tilde{k}, \tilde{\alpha})\tilde{y}
\end{aligned} \tag{4.24}$$

dimana :

$$\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{XA}(\tilde{k}, \tilde{\alpha}) + \mathbf{GB}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha}) \tag{4.25}$$

selanjutnya, pemilihan titik knok k optimal dan *bandwidth* $\tilde{\alpha}$ optimal menggunakan metode GCV. Sehingga nilai GCV dari model campuran *Spline truncated* dan Kernel diperoleh dengan cara :

$$GCV(k, \alpha) = \frac{MSE(k, \alpha)}{(n^{-1}tr(\mathbf{I} - \mathbf{M}(k, \alpha)))^2} \tag{4.26}$$

dimana :

$$MSE(k, \alpha) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

4.1.2 Perumusan Uji Hipotesis

Dilakukan pengujian hipotesis untuk parameter model yaitu parameter paramterik dan parameter nonparametrik Spline *Truncated*. Langkah pertama yang dilakukan untuk mendapatkan estimasi interval untuk parameter regresi semiparametrik campuran spline *truncated* dan kernel adalah mencari distribusi dari parameter-parameternya, yaitu $\tilde{\beta}$ adalah parameter komponen parametrik dan $\tilde{\theta}$ merupakan parameter komponen spline.

Komponen parametrik dan komponen nonparametrik spline dapat ditulis menjadi:

$$\mathbf{W}\tilde{\gamma}$$

dimana:

$$\mathbf{W} = [\mathbf{X} : \mathbf{G}]_{n \times ((p+1) + (1+r)q)}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix}_{(1+p) \times n}$$

$$\mathbf{G}(k) = \begin{bmatrix} t_{11} (t_{11} - K_{11})_+ \cdots (t_{11} - K_{1r})_+ \cdots t_{q1} (t_{q1} - K_{q1})_+ \cdots (t_{q1} - K_{qr})_+ \\ t_{12} (t_{12} - K_{11})_+ \cdots (t_{12} - K_{1r})_+ \cdots t_{q2} (t_{q2} - K_{q1})_+ \cdots (t_{q2} - K_{qr})_+ \\ \vdots \quad \vdots \quad \ddots \quad \vdots \quad \cdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\ t_{1n} (t_{1n} - K_{11})_+ \cdots (t_{1n} - K_{1r})_+ \cdots t_{qn} (t_{qn} - K_{q1})_+ \cdots (t_{qn} - K_{qr})_+ \end{bmatrix}_{(m+r)q \times n}$$

$$\tilde{\boldsymbol{\gamma}} = \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \cdots \\ \tilde{\boldsymbol{\theta}} \end{bmatrix}_{((1+p)+(m+r)q) \times 1}$$

sedangkan komponen nonparametrik kernel dapat ditulis menjadi:

$$\mathbf{D}(\tilde{\boldsymbol{\alpha}}) \tilde{\boldsymbol{y}}$$

dimana:

$$\mathbf{D}(\tilde{\boldsymbol{\alpha}}) = \mathbf{D}(\alpha_1) + \mathbf{D}(\alpha_2) + \cdots + \mathbf{D}(\alpha_r) \text{ dan } \tilde{\boldsymbol{y}} = [y_1 \quad y_2 \quad \cdots \quad y_n]^T.$$

Variabel respon $\tilde{\boldsymbol{y}}$ merupakan vektor berukuran $n \times 1$, matriks $\mathbf{W} = [\mathbf{X} : \mathbf{G}]$ adalah matriks yang memuat komponen parametrik dan komponen nonparametrik, \mathbf{X} adalah matriks yang memuat prediktor komponen parametrik yang berukuran $(1+p) \times n$ dan \mathbf{G} adalah matriks yang memuat komponen nonparametrik spline *truncated* berukuran $((m+r)q) \times 1$, $\tilde{\boldsymbol{\gamma}}$ merupakan vektor parameter dengan ukuran $((1+p)+(m+r)q) \times 1$. Pengujian hipotesis secara simultan diawali dengan menetapkan rumusan hipotesis yaitu:

$$H_0 : \mathbf{C}^T \tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau} \text{ lawan } H_1 : \mathbf{C}^T \tilde{\boldsymbol{\gamma}} \neq \boldsymbol{\tau}.$$

Estimator parameter dari model regresi semiparametrik campuran spline *truncated* dan kernel dapat diperoleh dengan menggunakan metode *Ordinary Least Square* (OLS). Berdasarkan model diberi hubungan:

$$\begin{aligned}
\tilde{y} &= \mathbf{W}\tilde{\gamma} + \mathbf{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon} \\
(\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} &= \mathbf{W}\tilde{\gamma} + \tilde{\varepsilon} \\
\tilde{y} &= (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^{-1} \mathbf{W}\tilde{\gamma} + (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^{-1} \tilde{\varepsilon} \\
&= \mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^*, \quad \tilde{\varepsilon}^* = (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^{-1} \tilde{\varepsilon}
\end{aligned} \tag{4.27}$$

selanjutnya diasumsikan $\tilde{\varepsilon}^* \sim N(\tilde{0}, \sigma_{\tilde{\varepsilon}^*}^2)$, dari persamaan 4.27 diatas dapat diperoleh persamaan sebagai berikut.

$$\begin{aligned}
\tilde{y} &= \mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^* \\
\tilde{\varepsilon}^* &= \tilde{y} - \mathbf{T}\tilde{\gamma}
\end{aligned} \tag{4.28}$$

selanjutnya dengan menjumlahkan kuadrat *error* bisa diperoleh hasil sebagai berikut.

$$\begin{aligned}
\sum_{i=1}^n \varepsilon_i^{*2} &= \tilde{\varepsilon}^{*T} \tilde{\varepsilon}^* = Q(\tilde{\gamma}) \\
\tilde{\varepsilon}^{*T} \tilde{\varepsilon}^* &= [\tilde{y} - \mathbf{T}\tilde{\gamma}]^T [\tilde{y} - \mathbf{T}\tilde{\gamma}] \\
&= [\tilde{y}^T - \tilde{\gamma}^T \mathbf{T}^T] [\tilde{y} - \mathbf{T}\tilde{\gamma}] \\
&= \tilde{y}^T \tilde{y} - \tilde{y}^T \mathbf{T}\tilde{\gamma} - \tilde{\gamma}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma} \\
&= \tilde{y}^T \tilde{y} - 2\tilde{\gamma}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma} \\
&= Q(\tilde{\gamma})
\end{aligned} \tag{4.29}$$

untuk mendapatkan estimator dari parameter, dilakukan derivatif parsial dari persamaan 4.29, $Q(\tilde{\gamma})$ terhadap $\tilde{\gamma}$ sebagai berikut.

$$\begin{aligned}
\frac{\partial Q(\tilde{\gamma})}{\partial \tilde{\gamma}} &= \frac{\tilde{y}^T \tilde{y} - 2\tilde{\gamma}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma}}{\partial \tilde{\gamma}} \\
&= -2\mathbf{T}^T \tilde{y} + 2\mathbf{T}^T \mathbf{T}\tilde{\gamma}
\end{aligned} \tag{4.30}$$

apabila persamaan 4.30 disamakan dengan nol, maka akan diperoleh persamaan berikut.

$$-2\mathbf{T}^T \tilde{y} + 2\mathbf{T}^T \mathbf{T}\hat{\tilde{\gamma}} = 0,$$

selanjutnya persamaan dapat ditulis dalam bentuk :

$$2\mathbf{T}^T \mathbf{T}\hat{\tilde{\gamma}} = 2\mathbf{T}^T \tilde{y},$$

sehingga diperoleh estimator parameter sebagai berikut.

$$\hat{\tilde{\gamma}} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y} \tag{4.31}$$

persamaan 4.34 dapat diringkas menjadi bentuk persamaan:

$$\hat{\gamma} = \mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y} \quad (4.32)$$

dimana :

$$\mathbf{M}(\tilde{k}, \tilde{\alpha}) = (\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \quad (4.33)$$

Setelah didapatkan estimator dari $\tilde{\gamma}$, langkah selanjutnya akan dicari distribusi dari $\hat{\gamma}$ dengan tahapan sebagai berikut. Apabila diasumsikan $\tilde{\varepsilon}^* \sim N(\tilde{0}, \sigma^2 \mathbf{I})$, karena \tilde{y} merupakan kombinasi linier dari $\tilde{\varepsilon}$ maka \tilde{y} juga mengikuti distribusi normal $\tilde{y} \sim N(E(\tilde{y}), Var(\tilde{y}))$. Selanjutnya akan dicari ekspektasi dan varians dari \tilde{y} yang diberikan sebagai berikut.

$$\begin{aligned} E(\tilde{y}) &= E(\mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^*) \\ &= \mathbf{T}\tilde{\gamma} + 0 \\ &= \mathbf{T}\tilde{\gamma} \end{aligned} \quad (4.34)$$

$$\begin{aligned} Var(\tilde{y}) &= Var(\mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^*) \\ &= 0 + Var(\tilde{\varepsilon}^*) \\ &= Var(\tilde{\varepsilon}^*) \\ &= \sigma^2 \mathbf{I} \end{aligned} \quad (4.35)$$

berdasarkan persamaan 4.34 dan 4.35 diperoleh $\tilde{y} \sim N(\mathbf{T}\tilde{\gamma}, \sigma^2 \mathbf{I})$. Karena sifat linieritas dari distribusi normal maka hal ini berlaku juga untuk $\hat{\gamma}$. Berikut diberikan ekspektasi dan varians dari $\hat{\gamma}$.

$$\begin{aligned} E(\hat{\gamma}) &= E(\mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y}) \\ &= \mathbf{M}(\tilde{k}, \tilde{\alpha})E(\tilde{y}) \\ &= (\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r E(\tilde{y}) \\ &= (\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \mathbf{T}\tilde{\gamma} \\ &= \tilde{\gamma} \end{aligned} \quad (4.36)$$

$$\begin{aligned}
\text{Var}(\hat{\tilde{\gamma}}) &= \text{Var}(\mathbf{M}(\tilde{k}, \tilde{\alpha})\tilde{y}) \\
&= ((\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)\text{Var}((\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)^T \\
&= \sigma^2\mathbf{I}(\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r\mathbf{T}(\mathbf{T}^r\mathbf{T})^{-1} \\
&= \sigma^2\mathbf{I}(\mathbf{T}^r\mathbf{T})^{-1}
\end{aligned} \tag{4.37}$$

$$\hat{\tilde{\gamma}} \sim N(\tilde{\gamma}, \sigma^2\mathbf{I}(\mathbf{T}^r\mathbf{T})^{-1}). \tag{4.38}$$

Selanjutnya adalah mencari bentuk distribusi dari $(\mathbf{C}^T\hat{\tilde{\gamma}}_\Omega - \boldsymbol{\tau})$ dengan uraian sebagai berikut:

$$\begin{aligned}
E(\mathbf{C}^T\hat{\tilde{\gamma}}_\Omega - \boldsymbol{\tau}) &= E((\mathbf{C}^T(\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)\tilde{y} - \boldsymbol{\tau}) \\
&= E(\mathbf{C}^T((\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r))E(\tilde{y}) - E(\boldsymbol{\tau}) \\
&= \mathbf{C}^T((\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)\tilde{\gamma} - \boldsymbol{\tau} \\
&= \mathbf{C}^T\tilde{\gamma} - \boldsymbol{\tau}
\end{aligned} \tag{4.39}$$

$$\begin{aligned}
\text{Var}(\mathbf{C}^T\hat{\tilde{\gamma}}_\Omega - \boldsymbol{\tau}) &= \text{Var}((\mathbf{C}^T(\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)\tilde{y} - \boldsymbol{\tau}) \\
&= \text{Var}((\mathbf{C}^T(\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)\tilde{y}) + \text{Var}(\boldsymbol{\tau}) \\
&= \mathbf{C}^T((\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r)\text{Var}(\tilde{y})\mathbf{T}(\mathbf{T}^r\mathbf{T})^{-1}\mathbf{C} \\
&= \sigma^2\mathbf{C}^T((\mathbf{T}^r\mathbf{T})^{-1}\mathbf{T}^r\mathbf{T}(\mathbf{T}^r\mathbf{T})^{-1})\mathbf{C} \\
&= \sigma^2\mathbf{I}(\mathbf{C}^T((\mathbf{T}^r\mathbf{T})^{-1})\mathbf{C})
\end{aligned} \tag{4.40}$$

karena $\hat{\tilde{\gamma}}_\Omega$ adalah fungsi linier dalam \tilde{y} , sehingga dapat dituliskan

$$(\mathbf{C}^T\hat{\tilde{\gamma}}_\Omega - \boldsymbol{\tau}) \sim N(\mathbf{C}^T\tilde{\gamma} - \boldsymbol{\tau}; \sigma^2\mathbf{I}(\mathbf{C}^T((\mathbf{T}^r\mathbf{T})^{-1})\mathbf{C})) \tag{4.41}$$

4.1.3 Pengujian Hipotesis

Untuk mengetahui apakah parameter $\tilde{\gamma}$ berpengaruh terhadap model maka dapat dilakukan uji hipotesis secara simultan. Pengujian hipotesis secara simultan diawali dengan menetapkan rumusan hipotesis yaitu

$$\begin{aligned}
H_0 : \mathbf{C}^T\tilde{\gamma} &= \boldsymbol{\tau} \\
H_1 : \mathbf{C}^T\tilde{\gamma} &\neq \boldsymbol{\tau}
\end{aligned} \tag{4.42}$$

dimana \mathbf{C}^T adalah suatu matriks yang berukuran $s \times d_1$ dan $\tilde{\gamma}$ adalah vektor berukuran $d_1 \times 1$ dengan $d_1 = (1 + p + (m+r)q) \times 1$ sehingga $\tilde{\gamma}$ berukuran $(1 + p + (m+r)q) \times 1$ dan adalah $\boldsymbol{\tau}$ vektor $s \times 1$ dengan elemennya berupa skalar.

Selanjutnya untuk mendapatkan statistik uji dari pengujian hipotesis simultan diselesaikan dengan menggunakan metode *Likelihood Ratio Test* (LRT). LRT digunakan untuk mencari rasio fungsi *likelihood* dari nilai maksimum dibawah ruang parameter terhadap nilai maksimum dibawah ruang parameter ω terhadap nilai maksimum dibawah ruang parameter Ω . Berdasarkan ruang parameter ω dan ruang parameter Ω maka dapat dicari fungsi *likelihood* dibawah Ω dan dibawah ω dengan langkah-langkah sebagai berikut.

a. Menentukan ruang parameter dibawah $H(\Omega)$ yaitu sebagai berikut.

$$\Omega = \left\{ \gamma = (\beta_0, \dots, \beta_p, \theta_{11}, \dots, \theta_{(r+m)1}, \dots, \theta_{1q}, \dots, \theta_{(m+r)q}), \sigma_\Omega^2 \right\}$$

b. Menentukan ruang parameter dibawah $H_0(\omega)$ yaitu sebagai berikut.

$$\omega = \left\{ \gamma = (\beta_0, \dots, \beta_p, \theta_{11}, \dots, \theta_{(r+m)1}, \dots, \theta_{1q}, \dots, \theta_{(m+r)q}), \sigma_\omega^2 \mid \mathbf{C}^T \tilde{\gamma} = \boldsymbol{\tau} \right\}$$

c. Menentukan fungsi *likelihood* dibawah ruang H.

Fungsi *likelihood* dibawah ruang $H(\Omega)$ adalah sebagai berikut. :

$$\begin{aligned} L(\tilde{\gamma}, \sigma^2) &= \prod_{i=1}^n \left((2\pi\sigma^2)^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (y_i - \mu(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}))^2 \right) \right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}))^2 \right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{T}\tilde{\gamma})^T (y_i - \mathbf{T}\tilde{\gamma}) \right) \end{aligned} \quad (4.43)$$

dengan $\tilde{y} = (y_1, y_2, \dots, y_n)^T$ dan $\tilde{\gamma} = (\beta_1, \beta_2, \dots, \beta_p, \theta_{11}, \theta_{12}, \dots, \theta_{q(m+r)})^T$.

Pertama diperhatikan ruang Ω , Fungsi *likelihood* diberikan oleh :

$$\begin{aligned}
L(\tilde{\gamma}_\Omega, \sigma_\Omega^2) &= (2\pi\sigma_\Omega^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_\Omega^2}(\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)\right) \\
&= (2\pi\sigma_\Omega^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_\Omega^2}(\tilde{y}^T - \tilde{\gamma}_\Omega^T \mathbf{T}^T)(\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)\right) \\
&= (2\pi\sigma_\Omega^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_\Omega^2}(\tilde{y}^T \tilde{y} - \tilde{y}^T \mathbf{T} \tilde{\gamma}_\Omega - \tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_\Omega^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega)\right) \\
&= (2\pi\sigma_\Omega^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_\Omega^2}(\tilde{y}^T \tilde{y} - 2(\tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y}) + \tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y} \mathbf{T} \tilde{\gamma}_\Omega)\right)
\end{aligned} \tag{4.44}$$

$$\begin{aligned}
\text{Log } L(\tilde{\gamma}_\Omega, \sigma_\Omega^2) &= \log\left((2\pi\sigma_\Omega^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_\Omega^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_\Omega^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega)\right)\right) \\
&= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_\Omega^2) - \frac{1}{2\sigma_\Omega^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_\Omega^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega)
\end{aligned} \tag{4.45}$$

$$\begin{aligned}
\frac{\partial \text{Log } L(\tilde{\gamma}_\Omega, \sigma_\Omega^2)}{\partial \tilde{\gamma}_\Omega} &= \frac{\partial}{\partial \tilde{\gamma}_\Omega} \left[-\frac{n}{2} \log(2\pi\sigma_\Omega^2) - \frac{1}{2\sigma_\Omega^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_\Omega^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega) \right] \\
&= 0 - \frac{1}{2\sigma_\Omega^2} (0 - 2\mathbf{T}^T \tilde{y} + 2\mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega)
\end{aligned} \tag{4.46}$$

$$\begin{aligned}
-2\mathbf{T}^T \tilde{y} + 2\mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega &= 0 \\
2\mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega &= 2\mathbf{T}^T \tilde{y} \\
\mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega &= \mathbf{T}^T \tilde{y} \\
\tilde{\gamma}_\Omega &= (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y}
\end{aligned} \tag{4.47}$$

$$\begin{aligned}
\frac{\partial L(\tilde{\gamma}_\Omega, \sigma_\Omega^2)}{\partial \sigma_\Omega^2} &= \frac{\partial}{\partial \sigma_\Omega^2} \left[-\frac{n}{2} \log(2\pi\sigma_\Omega^2) - \frac{1}{2\sigma_\Omega^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_\Omega^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_\Omega^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}_\Omega) \right] \\
&= -\frac{n}{2} \left(\frac{2\pi}{2\pi\sigma_\Omega^2} \right) + \frac{1}{2\sigma_\Omega^4} (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega) \\
&= -\frac{n}{2\sigma_\Omega^2} + \frac{1}{2\sigma_\Omega^4} (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega) \\
&= -n + \frac{1}{\sigma_\Omega^2} (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega) = 0
\end{aligned} \tag{4.48}$$

$$\begin{aligned}
n\sigma_\Omega^2 &= (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega) \\
\hat{\sigma}_\Omega^2 &= \frac{(\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)}{n}
\end{aligned}$$

akibatnya:

$$\begin{aligned}
Max_{\Omega} L(\tilde{\gamma}_{\Omega}, \sigma_{\Omega}^2) &= L(\hat{\Omega}) = L(\tilde{\gamma}_{\Omega}, \sigma_{\Omega}^2) \\
&= \left(2\pi\sigma_{\Omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\Omega}^2}(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})\right) \\
&= \left(2\pi\sigma_{\Omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})}{2\frac{(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})}{n}}\right) \\
&= \left(2\pi\sigma_{\Omega}^2\right)^{-\frac{n}{2}} e^{-\frac{n}{2}}
\end{aligned} \tag{4.49}$$

selanjutnya diperhatikan ruang ω . Fungsi *likelihood* diberikan oleh:

$$\begin{aligned}
L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2) &= \left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\omega}^2}(\tilde{y}^T - \tilde{\gamma}_{\omega}^T \mathbf{T}^T)(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})\right) \\
&= \left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\omega}^2}(\tilde{y}^T \tilde{y} - \tilde{y}^T \mathbf{T}\tilde{\gamma}_{\omega} - \tilde{\gamma}_{\omega}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_{\omega}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma}_{\omega})\right) \\
&= \left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\omega}^2}(\tilde{y}^T \tilde{y} - (\tilde{\gamma}_{\omega}^T \mathbf{T}^T \tilde{y})^T - \tilde{\gamma}_{\omega}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_{\omega}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma}_{\omega})\right)
\end{aligned} \tag{4.50}$$

$$\begin{aligned}
Log L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2) &= \log\left(\left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\omega}^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_{\omega}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_{\omega}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma}_{\omega})\right)\right) \\
&= -\frac{n}{2} \log(2\pi\sigma_{\omega}^2) - \frac{1}{2\sigma_{\omega}^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_{\omega}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_{\omega}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma}_{\omega})
\end{aligned} \tag{4.51}$$

$$\begin{aligned}
\frac{\partial Log L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2)}{\partial \sigma_{\omega}^2} &= \frac{\partial}{\partial \sigma_{\omega}^2} \left[-\frac{n}{2} \log(2\pi\sigma_{\omega}^2) - \frac{1}{2\sigma_{\omega}^2}(\tilde{y}^T \tilde{y} - 2\tilde{\gamma}_{\omega}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}_{\omega}^T \mathbf{T}^T \mathbf{T}\tilde{\gamma}_{\omega}) \right] \\
&= -\frac{n}{2} \left(\frac{2\pi}{2\pi\sigma_{\omega}^2} \right) + \frac{1}{2\sigma_{\omega}^4}(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}) \\
&= -\frac{n}{2\sigma_{\omega}^2} + \frac{1}{2\sigma_{\omega}^4}(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}) \\
&= -n + \frac{1}{\sigma_{\omega}^2}(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}) = 0
\end{aligned} \tag{4.52}$$

$$\begin{aligned}
n\sigma_{\omega}^2 &= (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}) \\
\sigma_{\omega}^2 &= \frac{(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})}{n}
\end{aligned}$$

akibatnya:

$$\begin{aligned}
\text{Max}_{\omega} L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2) &= \left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_{\omega}^2}(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})\right) \\
&= \left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} \exp\left(-\frac{(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})}{2\frac{(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})}{n}}\right) \\
&= \left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} e^{-\frac{n}{2}}
\end{aligned} \tag{4.53}$$

Statistik uji untuk hipotesis H_0 lawan H_1 diperoleh dengan metode

Likelihood Ratio Test (LRT) yaitu menyelesaikan rasio $L_{ratio} = \frac{\text{Max}_{\omega} L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2)}{\text{Max}_{\Omega} L(\tilde{\gamma}_{\Omega}, \sigma_{\Omega}^2)}$

dimana $\text{Max}_{\omega} L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2)$ adalah fungsi likelihood maksimum dibawah hipotesis dan

$\text{Max}_{\Omega} L(\tilde{\gamma}_{\Omega}, \sigma_{\Omega}^2)$ adalah fungsi likelihood maksimum dibawah populasi. Jika ingin menguji hipotesis H_0 lawan H_1 . Maka statistik uji untuk hipotesis ini adalah:

$$F = \frac{Q_1 / d_1}{Q_2 / d_2} \tag{4.54}$$

dimana

$$d_1 = (1 + p + (m + r)q) \times 1$$

$$d_2 = n - ((1 + p + (m + r)q) \times 1)$$

$$Q_1 = (\mathbf{C}^T \hat{\tilde{\gamma}}_{\Omega} - \boldsymbol{\tau})^T (\mathbf{C}^T ((\mathbf{T}^T \mathbf{T})^{-1}) \mathbf{C}) (\mathbf{C}^T \hat{\tilde{\gamma}}_{\Omega} - \boldsymbol{\tau})$$

$$Q_2 = (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})$$

$$\begin{aligned}
L_{ratio} &= \frac{\text{Max}_{\omega} L(\tilde{\gamma}_{\omega}, \sigma_{\omega}^2)}{\text{Max}_{\Omega} L(\tilde{\gamma}_{\Omega}, \sigma_{\Omega}^2)} = \frac{\left(2\pi\sigma_{\omega}^2\right)^{-\frac{n}{2}} e^{-\frac{n}{2}} \left(\sigma_{\omega}^2\right)^{-\frac{n}{2}}}{\left(2\pi\sigma_{\Omega}^2\right)^{-\frac{n}{2}} e^{-\frac{n}{2}} \left(\sigma_{\Omega}^2\right)^{-\frac{n}{2}}} = \left(\frac{\sigma_{\omega}^2}{\sigma_{\Omega}^2}\right)^{-\frac{n}{2}} = \left(\frac{\sigma_{\Omega}^2}{\sigma_{\omega}^2}\right)^{\frac{n}{2}} \\
&= \left(\frac{\left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}\right)^T \left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}\right)}{\frac{n}{\left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}\right)^T \left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}\right)}}\right)^{\frac{n}{2}} \\
&= \left(\frac{\left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}\right)^T \left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}\right)}{\left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}\right)^T \left(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}\right)}\right)^{\frac{n}{2}} \\
&= \left(\frac{Q_2}{A}\right)^{\frac{n}{2}} \tag{4.55}
\end{aligned}$$

dengan $Q_2 = (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})$ dan $Q_1 = (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})$.

Selanjutnya persamaan A akan disederhanakan dengan cara mensubstitusikan komponen Q_2 yang ditambahkan dengan komponen lain yaitu Q_1 , dimana komponen Q_1 memuat kendala yang sesuai rumusan hipotesis yang telah ditetapkan. Uraian komponen A diberikan sebagai berikut.

$$\begin{aligned}
A &= (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega})^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\omega}) \\
&= (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega} + \mathbf{T}\tilde{\gamma}_{\Omega} - \mathbf{T}\tilde{\gamma}_{\omega})^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega} + \mathbf{T}\tilde{\gamma}_{\Omega} - \mathbf{T}\tilde{\gamma}_{\omega}) \\
&= \left((\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}) + \mathbf{T}(\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})\right)^T \left((\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}) + \mathbf{T}(\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})\right) \\
&= \left((\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T + (\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})^T \mathbf{T}^T\right) \left((\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}) + \mathbf{T}(\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})\right) \\
&= \left((\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega}) + (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T \mathbf{T}(\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega}) + (\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})^T \mathbf{T}^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})\right) \\
&\quad + \left((\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})^T \mathbf{T}^T\right) \mathbf{T}(\tilde{\gamma}_{\Omega} - \tilde{\gamma}_{\omega})
\end{aligned} \tag{4.56}$$

dalam persamaan 4.56 komponen vektor $(\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})^T \mathbf{T}$ dan vektor $\mathbf{T}^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_{\Omega})$ keduanya bernilai nol. Hal ini dapat dilihat pada uraian berikut ini:

$$\hat{\tilde{\gamma}}_{\Omega} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y} \tag{4.57}$$

dapat diuraikan menjadi seperti berikut.

$$\begin{aligned} (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T \mathbf{T} &= \left(\tilde{y} - \mathbf{T} \left((\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y} \right)^T \mathbf{T} \right) \\ &= \left(\tilde{y}^T \mathbf{T} - \tilde{y}^T \mathbf{T} \left((\mathbf{T}^T \mathbf{T})^{-1} (\mathbf{T}^T \mathbf{T}) \right) \right) \end{aligned} \quad (4.58)$$

dalam persamaan 4.58 terdapat komponen matriks identitas yaitu $(\mathbf{T}^T \mathbf{T})^{-1} (\mathbf{T}^T \mathbf{T}) = \mathbf{I}$ sehingga persamaan 4.58 menjadi seperti berikut.

$$\tilde{y}^T \mathbf{T} - \tilde{y}^T \mathbf{T} = \mathbf{0}. \quad (4.59)$$

Selanjutnya vektor $\mathbf{T}^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)$ dapat diuraikan menjadi seperti berikut.

$$\begin{aligned} \mathbf{T}^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega) &= \mathbf{T}^T \left(\tilde{y} - \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y} \right) \\ &= \mathbf{T}^T \tilde{y} - \mathbf{T}^T \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y} \end{aligned} \quad (4.60)$$

dari persamaan 4.60 memuat komponen matriks identitas yaitu $(\mathbf{T}^T \mathbf{T})^{-1} (\mathbf{T}^T \mathbf{T}) = \mathbf{I}$.

Sehingga persamaan 4.60 dapat dituliskan menjadi seperti berikut.

$$\mathbf{T}^T \tilde{y} - \mathbf{T}^T \tilde{y} = \mathbf{0} \quad (4.61)$$

dari persamaan 4.59 dan 4.61 terbukti bahwa komponen vektor $(\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T \mathbf{T}$ dan vektor $\mathbf{T}^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)$ keduanya bernilai nol. Sehingga persamaan 4.56 dapat ditulis sebagai berikut.

$$\mathbf{A} = (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\tilde{\gamma}_\Omega) + (\tilde{\gamma}_\Omega - \tilde{\gamma}_\omega)^T \mathbf{T}^T \mathbf{T} (\tilde{\gamma}_\Omega - \tilde{\gamma}_\omega). \quad (4.62)$$

Selanjutnya mensubstitusikan persamaan 4.63 kedalam persamaan 4.64

$$\hat{\gamma}_\Omega - \hat{\gamma}_\omega = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right) (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau}) \quad (4.63)$$

$$\begin{aligned} \mathbf{A} &= (\tilde{y} - \mathbf{T}\hat{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\hat{\gamma}_\Omega) + (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau})^T \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \right) \times \\ &\quad \left(\mathbf{T}^T \mathbf{T} \right) (\mathbf{T}^T \mathbf{T})^{-1} \left(\mathbf{C} \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau}) \right) \\ &= (\tilde{y} - \mathbf{T}\hat{\gamma}_\Omega)^T (\tilde{y} - \mathbf{T}\hat{\gamma}_\Omega) + (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau})^T \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau}) \\ &= \mathcal{Q}_1 + \mathcal{Q}_2 \end{aligned} \quad (4.64)$$

dimana

$$\mathbf{Q}_1 = (\mathbf{C}^T \hat{\boldsymbol{\gamma}}_{\Omega} - \boldsymbol{\tau})^T \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} (\mathbf{C}^T \hat{\boldsymbol{\gamma}}_{\Omega} - \boldsymbol{\tau}) \quad (4.65)$$

dengan menggunakan persamaan 4.61 selanjutnya dapat dituliskan sebagai berikut.

$$L_{\text{ratio}} = \left(\frac{\mathbf{Q}_2}{\mathbf{Q}_2 + \mathbf{Q}_1} \right)^{\frac{n}{2}} = \left(\frac{1}{1 + \frac{\mathbf{Q}_1}{\mathbf{Q}_2}} \right)^{\frac{n}{2}} = \left(1 + \frac{\mathbf{Q}_1}{\mathbf{Q}_2} \right)^{-\frac{n}{2}} \quad (4.66)$$

L_{ratio} adalah fungsi dari $\frac{\mathbf{Q}_1}{\mathbf{Q}_2}$, dimana L_{ratio} akan bernilai kecil untuk $\frac{\mathbf{Q}_1}{\mathbf{Q}_2}$ besar,

sehingga $\frac{\mathbf{Q}_1}{\mathbf{Q}_2}$ di jadikan sebagai statistik uji untuk L_{ratio} .

Berdasarkan Persamaan 4.67 ini, dan sedikit penjabaran didapatkan statistik uji untuk uji hipotesis $H_0 : \mathbf{C}^T \tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau}$ lawan $H_0 : \mathbf{C}^T \tilde{\boldsymbol{\gamma}} \neq \boldsymbol{\tau}$ dari model adalah:

$$\mathbf{F} = \frac{\mathbf{Q}_1 / d_1}{\mathbf{Q}_2 / d_2} .$$

4.1.4 Mendapatkan Distribusi Statistik Uji

Jika \mathbf{Q}_1 dan \mathbf{Q}_2 masing-masing diberikan dalam Persamaan 4.65 dan 4.55 maka statistik

$$\mathbf{F} = \frac{\mathbf{Q}_1 / d_1}{\mathbf{Q}_2 / d_2} \sim F_{(d_1, d_2)} .$$

Selanjutnya menunjukkan bahwa $\mathbf{Q}_1 / \sigma^2 \sim \chi^2_{(d_1)}$, menunjukkan bahwa $\mathbf{Q}_2 / \sigma^2 \sim \chi^2_{(d_2)}$, serta menunjukkan bahwa \mathbf{Q}_1 dan \mathbf{Q}_2 independen.

Sebagai langkah awal adalah menuliskan kembali \mathbf{Q}_1 , yaitu

$$\mathbf{Q}_1 = (\mathbf{C}^T \hat{\boldsymbol{\gamma}}_{\Omega} - \boldsymbol{\tau})^T \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} (\mathbf{C}^T \hat{\boldsymbol{\gamma}}_{\Omega} - \boldsymbol{\tau})$$

Dengan menulis bentuk \mathbf{Q}_1 yang dibagi dengan σ^2 didapatkan

$$\frac{\mathbf{Q}_1}{\sigma^2} = (\mathbf{C}^T \hat{\boldsymbol{\gamma}}_{\Omega} - \boldsymbol{\tau})^T \frac{\left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1}}{\sigma^2} (\mathbf{C}^T \hat{\boldsymbol{\gamma}}_{\Omega} - \boldsymbol{\tau})$$

Selanjutnya untuk

$$\mathbf{A} = \frac{\left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C}\right)^{-1}}{\sigma^2}$$

dan

$$\mathbf{V} = \sigma^2 \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C}\right)$$

dapat dituliskan bahwa komponen matriks $(\mathbf{A}\mathbf{V})^T (\mathbf{A}\mathbf{V}) = (\mathbf{A}\mathbf{V})$ adalah suatu matriks idempoten. Dapat dituliskan bentuk distribusi dari $\frac{\mathbf{Q}_1}{\sigma^2}$ yang dituliskan dalam bentuk

$$\frac{\mathbf{Q}_1}{\sigma^2} \sim \chi^2_{\left\{d_1, (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau})^T \frac{(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C})^{-1}}{\sigma^2} (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau})\right\}}$$

atau dalam bentuk sebagai berikut:

$$\frac{\mathbf{Q}_1}{\sigma^2} \sim \chi^2_{(\text{rank}(\mathbf{A}), \mathbf{D})}$$

dimana

$$\mathbf{D} = \mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau})^T \frac{(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C})^{-1}}{\sigma^2} (\mathbf{C}^T \hat{\gamma}_\Omega - \boldsymbol{\tau})$$

Karena \mathbf{A} adalah matriks *full rank* dengan ukuran d_1 dan di bawah hipotesis H_0 , matriks $\mathbf{D} = \mathbf{0}$ sehingga bentuk dalam Persamaannya menjadi

$$\frac{\mathbf{Q}_1}{\sigma^2} \sim \chi^2_{(d_1)}$$

Kemudian untuk menunjukkan $\mathbf{Q}_2 / \sigma^2 \sim \chi^2_{(d_2)}$ dilakukan dengan memodifikasi \mathbf{Q}_2 ke dalam bentuk *quadratic form* dengan vektor *quadratic form* adalah $\tilde{\mathbf{y}}$. Memodifikasi \mathbf{Q}_2 ke dalam bentuk *quadratic form* diberikan dalam uraian sebagai berikut

$$\begin{aligned} \mathbf{Q}_2 &= (\tilde{\mathbf{y}} - \mathbf{T} \tilde{\gamma}_\Omega)^T (\tilde{\mathbf{y}} - \mathbf{T} \tilde{\gamma}_\Omega) \\ &= \tilde{\mathbf{y}}^T (\mathbf{J}\mathbf{J}) \tilde{\mathbf{y}} \end{aligned} \tag{4.67}$$

dimana

$$\mathbf{J} = \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \right)$$

$\mathbf{J}' \mathbf{J} = \mathbf{J}$ adalah matriks yang simetris dan idempoten, sehingga bentuk pada Persamaan 4.67 dapat dituliskan menjadi

$$\mathbf{Q}_2 = \tilde{\mathbf{y}}^T \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \right) \tilde{\mathbf{y}}$$

dimana

$$\tilde{\mathbf{y}} \sim N(\mathbf{T}\tilde{\boldsymbol{\gamma}}, \sigma^2 \mathbf{I})$$

serta

$$\mathbf{A}_* = \frac{\mathbf{A}_{**}}{\sigma^2}$$

dengan

$$\mathbf{A}_{**} = \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \right)$$

dan

$$\mathbf{V}_* = \sigma^2 \mathbf{I}.$$

diketahui bahwa $(\mathbf{A}_* \mathbf{V}_*)^T (\mathbf{A}_* \mathbf{V}_*) = (\mathbf{A}_* \mathbf{V}_*)$ adalah matriks idempoten.

Selanjutnya dapat dituliskan sebagai berikut

$$\frac{\mathbf{Q}_2}{\sigma^2} = \tilde{\mathbf{y}}^T \frac{\mathbf{A}_{**}}{\sigma^2} \tilde{\mathbf{y}}$$

setelah itu didapatkan

$$\frac{\mathbf{Q}_2}{\sigma^2} \sim \chi^2_{(\text{rank}(\mathbf{A}_*), \mathbf{E})}$$

dengan

$$\mathbf{E} = \frac{\mathbf{E}_*}{2\sigma^2}$$

dimana

$$\mathbf{E}_* = (\mathbf{T}\tilde{\boldsymbol{\gamma}}_\Omega)^T \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T}) \right) \mathbf{T}^T \mathbf{T}\tilde{\boldsymbol{\gamma}}_\Omega$$

didapatkan bahwa

$$\text{rank}(\mathbf{A}_*) = \text{rank}(\mathbf{A}_* \mathbf{V}_*) = \text{tr}(\mathbf{A} \mathbf{V}) = n - (1 + p + (m + r)q) \times 1.$$

Dari penjabaran matriks \mathbf{E} didapatkan hasilnya adalah nol, sehingga Persamaannya dapat dituliskan menjadi:

$$\frac{\mathbf{Q}_2}{\sigma^2} = \sim \chi^2_{(d_2)}.$$

Selanjutnya untuk menunjukkan apakah \mathbf{Q}_1 dan \mathbf{Q}_2 saling independen atau tidak, adalah dengan menunjukkan bahwa $\mathbf{GVF} = \mathbf{0}$, dimana $\mathbf{V} = \sigma^2 \mathbf{I}$ adalah varians dari vektor *quadratic form* \mathbf{Q}_1 dan \mathbf{Q}_2 , \mathbf{G} dan \mathbf{F} masing-masing adalah matriks *quadratic form* dari \mathbf{Q}_1 dan \mathbf{Q}_2 . Dengan memodifikasi \mathbf{Q}_1 dan \mathbf{Q}_2 ke dalam *quadratic form* dengan vektor variabel random yang berdistribusi normal didapatkan

$$\mathbf{Q}_1 = \left(\tilde{\mathbf{y}} - \mathbf{TC}(\mathbf{C}^T \mathbf{C})^{-1} \boldsymbol{\tau} \right)^T \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \left(\mathbf{C}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} \mathbf{C}^T \left((\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T} \left(\tilde{\mathbf{y}} - \mathbf{TC}(\mathbf{C}^T \mathbf{C})^{-1} \boldsymbol{\tau} \right) \right)$$

dan

$$\mathbf{Q}_2 = \left(\tilde{\mathbf{y}} - \mathbf{TC}[\mathbf{C}^T \mathbf{C}]^{-1} \boldsymbol{\tau} \right)^T \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \right) \left(\tilde{\mathbf{y}} - \mathbf{TC}(\mathbf{C}^T \mathbf{C}) \right)$$

diketahui bahwa \mathbf{Q}_1 dan \mathbf{Q}_2 keduanya adalah *quadratic form* dalam vektor $\mathbf{x} = \left(\tilde{\mathbf{y}} - \mathbf{TC}(\mathbf{C}^T \mathbf{C}) \right)$ yang berdistribusi normal, dengan masing-masing matriks *quadratic form* untuk \mathbf{Q}_1 adalah

$$\mathbf{H} = \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \left(\mathbf{C}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{C} \right)^{-1} \mathbf{C}^T \left((\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \right) \quad (4.68)$$

dan matriks *quadratic form* untuk \mathbf{Q}_2 adalah

$$\mathbf{G} = \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \right) \quad (4.69)$$

dengan menggunakan hasil kali matriks persamaan 4.68 dan 4.69 serta varians dari $\left(\tilde{\mathbf{y}} - \mathbf{TC}(\mathbf{C}^T \mathbf{C})^{-1} \boldsymbol{\tau} \right)$ dengan sedikit penjabaran didapatkan

$$\begin{aligned} \mathbf{GVH} &= \mathbf{G}(\sigma^2 \mathbf{I})\mathbf{H} \\ &= \sigma^2 \mathbf{G}\mathbf{I}\mathbf{H} \\ &= \mathbf{0}. \end{aligned} \quad (4.70)$$

Didapatkan bentuk distribusi dari statistik uji untuk uji hipotesis $H_0 : \mathbf{C}'\tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau}$ lawan

$H_1 : \mathbf{C}'\tilde{\boldsymbol{\gamma}} \neq \boldsymbol{\tau}$ adalah

$$\mathbf{F} = \frac{\mathbf{Q}_1 / d_1}{\mathbf{Q}_2 / d_2} \sim F_{(d_1, d_2)}.$$

daerah kritis dari uji simultan diperoleh dari menyelesaikan persamaan

$$P(\mathbf{F} > k^* | \mathbf{C}\tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau}) = \alpha$$

dimana α adalah tingkat signifikansi dan k^* adalah suatu konstanta tertentu yang nilainya bergantung pada α .

Daerah penolakan hipotesis (daerah kritis) untuk menguji hipotesis $H_0 : \mathbf{C}'\tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau}$

lawan $H_1 : \mathbf{C}'\tilde{\boldsymbol{\gamma}} \neq \boldsymbol{\tau}$, adalah himpunan semua titik yang memenuhi

$$L_{\text{ratio}} = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k$$

dengan menjabarkan ketaksamaan didapatkan

$$\mathbf{F} > k^*$$

dimana

$$k^* = \frac{d_2}{d_1} \left(k^{\frac{-2}{n}} - 1 \right).$$

Selanjutnya jika diberikan tingkat signifikansi α maka k^* diperoleh dari persamaan

$$\alpha = P(\mathbf{F} > k^* | \mathbf{C}\tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau}),$$

dimana $\mathbf{F} \sim F_{(d_1, d_2)}$.

Dengan hipotesis $H_0 : \mathbf{C}'\tilde{\boldsymbol{\gamma}} = \boldsymbol{\tau}$ lawan $H_1 : \mathbf{C}'\boldsymbol{\gamma} \neq \boldsymbol{\tau}$, hipotesis H_0 ditolak ketika nilai

dari $L_{\text{ratio}} = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k$, dengan uraian sebagai berikut

$$L_{\text{ratio}} = \left(\frac{1}{1 + \frac{\mathbf{Q}_1}{\mathbf{Q}_2}} \right)^{\frac{n}{2}} < k, \quad 0 < k < 1$$

$$\begin{aligned} &\Rightarrow \left(\frac{1}{1 + \frac{Q_1}{Q_2}} \right) < k^{\frac{2}{n}} \\ &\Rightarrow \left(\frac{1}{k^{\frac{2}{n}}} \right) < \left(1 + \frac{Q_1}{Q_2} \right) \\ &\Rightarrow \left(\frac{1}{k^{\frac{2}{n}}} \right) - 1 < \left(\frac{Q_1}{Q_2} \right) \\ &\Rightarrow \left(\frac{Q_1}{Q_2} \right) > k^* \end{aligned}$$

dengan k^* adalah suatu nilai yang didapat dari taraf signifikansi α serta

$$\begin{aligned} \alpha &= P(\text{menolak } H_0 | H_0 \text{ benar}) \\ &= P(\mathbf{W} > k^* | \mathbf{C}\boldsymbol{\gamma} = \boldsymbol{\tau}), \mathbf{F} \sim F_{(\alpha; s; ((1+p+(m+r)q))} \end{aligned}$$

atau dengan kata lain bahwa dalam Hipotesis H_0 akan ditolak jika $\mathbf{F} > F_{(\alpha; s; 1+(p+(m+r)q))}$.

4.2 Aplikasi pada Data TFR di Provinsi Jawa Timur

Model dan estimasi interval regresi semiparametrik Spline *truncated* yang telah dibahas pada sub bab 4.1 akan diaplikasikan pada data TFR Tahun 2015 di Provinsi Jawa Timur. Variabel respon yang digunakan adalah TFR (\mathbf{Y}) sedangkan variabel prediktor yang digunakan sebanyak 4 variabel yaitu Indeks *Unmet Need* (\mathbf{X}_1), *Age Spesific Fertility Rate* (\mathbf{T}_1), Indeks Pembangunan Manusia (\mathbf{T}_2), dan Angka Kematian Bayi (\mathbf{Z}_1). Unit observasi yang digunakan sebanyak 38 kabupaten/kota.

4.2.1 Analisis Deskriptif

Hasil dari statistik deskriptif dapat digunakan untuk inisiasi titik-titik knot pada tahap analisis selanjutnya.

Tabel 4.1 Statistik Deskriptif Variabel Respon dan Variabel Prediktor

| Variabel | Mean | Minimum | Maksimum | Standar Deviasi | Range |
|----------|-------|---------|----------|-----------------|-------|
| Y | 2,06 | 1,52 | 2,45 | 0,21 | 0,93 |
| X_1 | 69,11 | 58,18 | 80,05 | 5,40 | 21,87 |
| T_1 | 16,01 | 6,87 | 31,77 | 5,01 | 24,9 |
| T_2 | 36,21 | 6,40 | 87,00 | 18,79 | 80,60 |
| Z_1 | 30,92 | 17,27 | 60,51 | 12,09 | 43,24 |

Dari tabel 4.1 di atas dapat dilihat karakteristik masing- masing variabel, baik variabel respon maupun variabel prediktor. Berdasarkan tersebut di atas variabel respon (Y) yaitu TFR di Provinsi Jawa Timur memiliki nilai rata-rata TFR pada Tahun 2015 sebesar 2,06 atau bisa dikatakan kalau rata-ratanya lebih mendekati ke 2 orang dengan standar deviasi sebesar 0,21. Angka TFR tertinggi sebesar 2,45 terdapat pada Kabupaten Sampang. Sedangkan Angka TFR terendah dengan nilai 2,05 terdapat pada Kabupaten Sumenep. Range nilai TFR Provinsi Jawa Timur dengan 38 kabupaten/kota sebesar 0,93.

Berikut adalah gambaran karakteristik untuk masing-masing variabel prediktor, yaitu *Unmet Need* (X_1), *Age Spesific Fertility Rate* (T_1), Indeks Pembangunan Manusia (T_2) dan Angka Kematian Bayi (Z_1).

- Rata-rata *Unmet Need* di Provinsi Jawa Timur pada Tahun 2015 adalah 16,01 dengan standar deviasi sebesar 5,01. *Unmet Need* tertinggi pada Kabupaten Bangkalan sebesar 31,77 dan yang paling terendah ada di Kabupaten Bondowoso dengan angka 6,87 dengan range sebesar 24,90.
- Rata-rata ASFR di Provinsi Jawa Timur pada Tahun 2015 adalah 36,21 dengan standar deviasi sebesar 18,79. ASFR tertinggi pada Kabupaten Bondowoso sebesar 87,00 dan yang paling terendah ada di Kota Malang dengan angka 6,40.
- Rata-rata Indeks Pembangunan Manusia di Provinsi Jawa Timur pada Tahun 2015 adalah 69,11 dengan standar deviasi sebesar 5,4. Indeks Pembangunan Manusia tertinggi pada Kota Malang sebesar 80,05 dan paling terendah ada di Kabupaten Sampang dengan angka 58,18 dengan range sebesar 24,90.

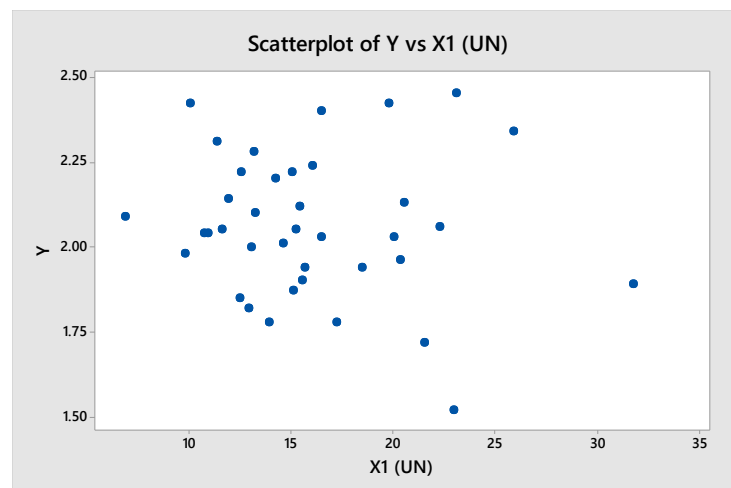
- d. Rata-rata AKB di Provinsi Jawa Timur pada Tahun 2015 adalah 30,92 atau 31 orang per 1000 kelahiran hidup dengan standar deviasi sebesar 12,09. AKB tertinggi pada Kabupaten Probolinggo sebesar 60,51 atau 61 orang dan yang paling terendah ada di Kota Blitar dengan angka 17,27 atau 17 orang dengan range sebesar 43,24.

4.2.2 Pemodelan TFR Menggunakan Regresi Semiparametrik Campuran Spline *Truncated* dan *Kernel*

Eksplorasi data yang dilakukan untuk mengidentifikasi bentuk pola hubungan antara variabel respon dengan masing-masing variabel prediktor adalah melalui *scatter plot*. Selanjutnya dapat digunakan untuk menentukan variabel sebagai komponen parametrik dan komponen nonparametrik.

4.2.2.1 Penentuan Variabel Komponen Parametrik dan Komponen Nonparametrik

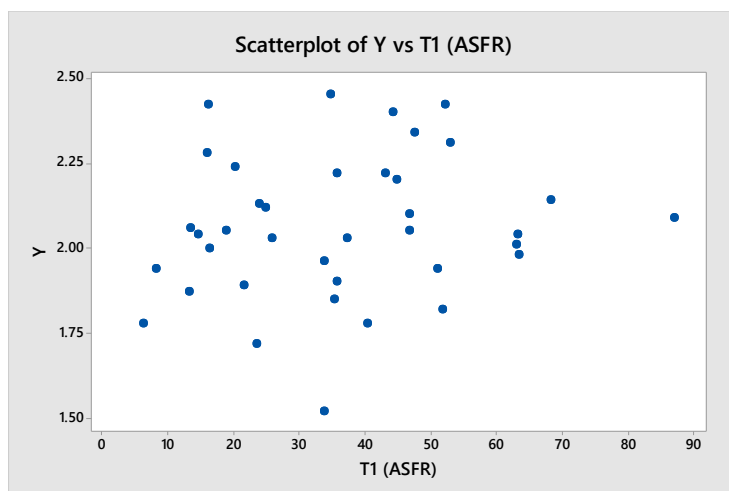
Tahap awal yang dilakukan untuk melakukan proses pemodelan regresi semiparametrik Spline *truncated* linier adalah dengan membuat *scatter plot* antara variabel respon dengan masing-masing variabel prediktor. Dari hasil *scatter plot* ini dapat dilihat bentuk pola hubungan antara variabel respon dengan masing-masing variabel prediktor. Berikut adalah hasil *scatter plot* antara variabel respon dengan masing-masing variabel prediktor:



Gambar 4.1 Plot antara TFR dengan *Unmet Need*

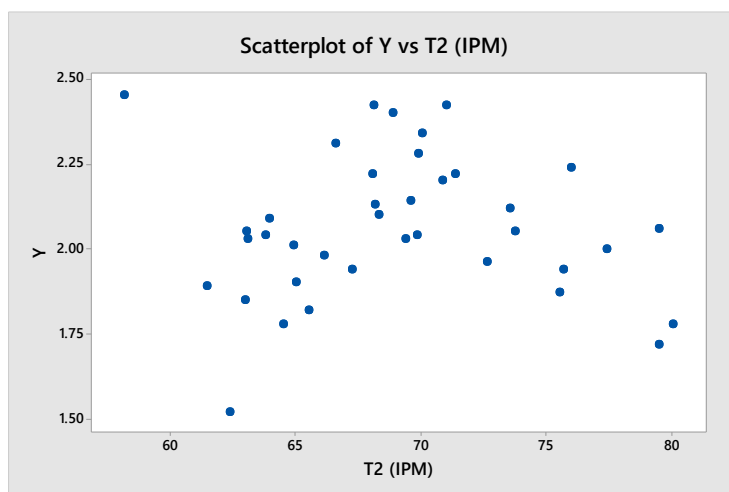
Gambar 4.1 menunjukkan bentuk pola hubungan antara variabel *Unmet Need* dengan TFR yang cenderung membentuk pola tertentu. Pola hubungan yang

terlihat cenderung mengikuti atau linier, selain dari hasil tersebut *Unmet Need* menghasilkan GCV paling minimum saat diposisikan sebagai variabel parametrik setelah sebelumnya dicoba sebagai variabel nonparametrik. Dari pertimbangan tersebut maka variabel *Unmet Need* digunakan sebagai komponen parametrik.



Gambar 4.2 Scatter Plot antara TFR dengan ASFR

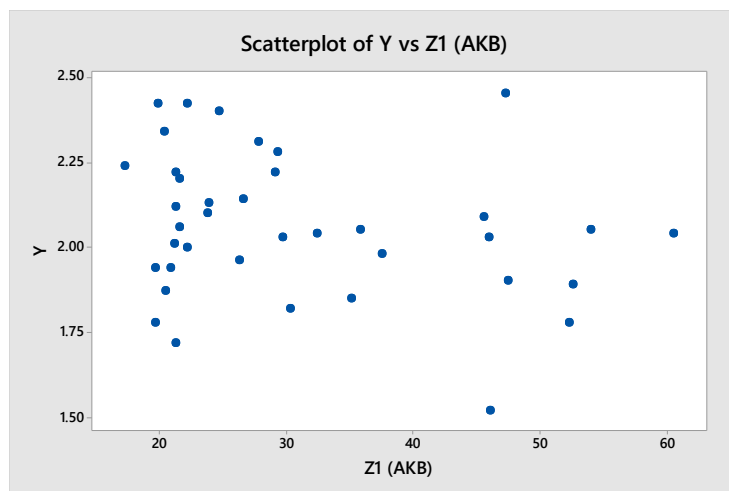
Gambar 4.2 menunjukkan bentuk pola hubungan antara variabel TFR dengan ASFR yang cenderung tidak membentuk pola tertentu. Pola hubungan yang terlihat cenderung mengalami perubahan perilaku pada sub interval tertentu. Dari hasil *scatter plot* diatas variabel jumlah ASFR digunakan sebagai komponen nonparametrik.



Gambar 4.3 Scatter plot antara TFR dengan IPM

Gambar 4.3 menunjukkan bentuk pola hubungan antara variabel TFR dengan IPM yang cenderung tidak membentuk pola tertentu. Pola hubungan yang

terlihat cenderung mengalami perubahan perilaku pada sub interval tertentu. Dari hasil *scatter plot* diatas variabel IPM digunakan sebagai komponen nonparametrik.



Gambar 4.4 *Scatter Plot* antara TFR dengan AKB

Gambar 4.4 menunjukkan bentuk pola hubungan antara variabel TFR dengan AKB yang cenderung tidak membentuk pola tertentu sehingga dimodelkan dengan secara nonparametrik *kernel*. Dari hasil kombinasi *scatter plot* dapat disimpulkan bahwa variabel AKB digunakan sebagai komponen nonparametrik *kernel*.

Ringkasan hasil penentuan komponen parametrik dan komponen nonparametrik disajikan pada Tabel 4.2.

Tabel 4.2 Komponen Parametrik dan Nonparametrik

| Notasi | Variabel | Komponen |
|--------|---------------------|---------------|
| X_1 | Unmet Need | Parametrik |
| T_1 | ASFR | Nonparametrik |
| T_2 | IPM | Nonparametrik |
| Z_1 | Angka Kematian Bayi | Nonparametrik |

4.2.3 Model Regresi Semiparametrik Spline *Truncated* Linier

Berdasarkan *scatter plot* antara TFR dengan variabel-variabel yang diduga mempengaruhinya, maka metode yang akan digunakan dalam analisis ini adalah analisis regresi semiparametrik campuran Spline *truncated* dan *kernel*.

4.2.3.1 Memodelkan TFR di Provinsi Jawa Timur Menggunakan Semiparametrik Campuran Spline *Truncated* dan Kernel Satu Knot

Model regresi semiparametrik Spline *truncated* linier dengan satu variabel komponen parametrik dan empat variabel komponen nonparametrik dengan satu titik knot adalah sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + h(z_i) + \varepsilon_i ; i = 1, 2, \dots, 38$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan satu knot disajikan dalam Tabel 4.3 berikut.

Tabel 4.3 Perbandingan Nilai GCV dengan menggunakan Satu Titik Knot

| Spline | | Kernel | GCV | R ² |
|----------------------------------|----------------------------------|-----------|----------|----------------|
| Knot | | Bandwidth | | |
| t ₁ = K ₁₁ | t ₂ = K ₂₁ | α | | |
| 56,3 | 71,72 | 0,0485 | 0,004164 | 95,06 |
| 60,13 | 72,76 | 0,0494 | 0,004268 | 94,69 |
| 52,46 | 70,68 | 0,048 | 0,004301 | 94,99 |
| 63,97 | 73,8 | 0,0501 | 0,00445 | 94,28 |

Berdasarkan Tabel 4.3 nilai GCV minimum yang dihasilkan adalah sebesar 0,004164. Lokasi titik knot pada variabel (t₁) yaitu 56,3(K₁₁) dan (t₂) yaitu 71,72 (K₁₂) sedangkan bandwidth yang diberikan sebesar α = 0,0485.

4.2.3.2 Memodelkan TFR di Provinsi Jawa Timur dengan Komponen Spline *Truncated* Dua Titik Knot

Model regresi semiparametrik spline *truncated* dengan menggunakan dua titik knot dengan prediktor satu komponen parametrik dan lima komponen nonparametrik adalah sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{31} (t_{1i} - K_{21})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{32} (t_{2i} - K_{22})_+^1 + h(z_i) + \varepsilon_i ; i = 1, 2, \dots, 38$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan satu knot disajikan dalam Tabel 4.4

Tabel 4.4 Perbandingan Nilai GCV dengan menggunakan Dua Titik Knot

| Spline | | Kernel | | GCV | R ² |
|----------------------------------|----------------------------------|-----------|--|----------|----------------|
| Knot | | Bandwidth | | | |
| t ₁ = K ₁₁ | t ₁ = K ₁₂ | α | | | |
| t ₂ = K ₂₁ | t ₂ = K ₂₂ | | | | |
| 68,4 | 80,8 | 0,04472 | | 0,004185 | 96,32 |
| 75 | 78,37 | | | | |
| 18,8 | 56 | 0,04909 | | 0,004207 | 95,32 |
| 61,54 | 71,64 | | | | |
| 68,4 | 74,6 | 0,04553 | | 0,004303 | 95,91 |
| 75 | 76,69 | | | | |
| 12,6 | 56 | 0,05001 | | 0,004346 | 95,09 |
| 59,86 | 71,64 | | | | |

Berdasarkan Tabel 4.4 nilai GCV minimum yang dihasilkan adalah sebesar 0,004185. Lokasi titik knot pada variabel (t₁) yaitu 68,4 (K₁₁), 80,8 (K₁₂), untuk (t₂) yaitu 75 (K₂₁) 78,37 (K₂₂) sedangkan bandwidth yang diberikan sebesar α = 0,04472.

4.2.3.3 Memodelkan TFR di Provinsi Jawa Timur dengan Komponen Spline Truncated Tiga Titik Knot

Model regresi semiparametrik spline *truncated* dengan menggunakan tiga titik knot dengan prediktor satu komponen parametrik dan lima komponen nonparametrik adalah sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{31} (t_{1i} - K_{21})_+^1 + \theta_{41} (t_{1i} - K_{21})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{32} (t_{2i} - K_{22})_+^1 + \theta_{42} (t_{2i} - K_{32})_+^1 + h(z_i) + \varepsilon_i ; i = 1, 2, \dots, 38$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan tiga knot disajikan dalam Tabel 4.5

Tabel 4.5 Perbandingan Nilai GCV dengan menggunakan Tiga Titik Knot

| SplineKnot | | Kernel | GCV | R ² |
|----------------------------------|----------------------------------|-----------|----------|----------------|
| Knot | | Bandwidth | | |
| t ₁ = K ₁₁ | t ₁ = K ₁₂ | α | | |
| t ₂ = K ₂₁ | t ₂ = K ₂₂ | | | |
| t ₃ = K ₃₁ | t ₃ = K ₃₂ | | | |
| 57,69 | 72,1 | 0,04277 | 0,004115 | 97,05 |
| 72,35 | 76,07 | | | |
| 79,67 | 78,06 | | | |
| 65,02 | 74,09 | 0,04357 | 0,00423 | 96,66 |
| 72,35 | 76,07 | | | |
| 79,67 | 78,06 | | | |
| 50,36 | 70,11 | 0,04242 | 0,004295 | 97,05 |
| 72,35 | 76,07 | | | |
| 79,67 | 78,06 | | | |
| 43,04 | 68,12 | 0,04213 | 0,004412 | 97,12 |
| 72,35 | 76,07 | | | |
| 79,67 | 78,06 | | | |

Berdasarkan Tabel 4.5 nilai GCV minimum yang dihasilkan adalah sebesar 0,004115 Lokasi titik knot pada variabel (t₁) yaitu 57,69 (K₁₁) 72,35(K₁₂) 79,67(K₁₃) untuk (t₂) yaitu 72,1(K₂₁) 76,07(K₂₂) 78,06(K₂₃) sedangkan bandwidth yang diberikan sebesar α = 0,04277.

4.2.3.4 Memodelkan TFR di Provinsi Jawa Timur dengan Komponen Spline Truncated Kombinasi Titik Knot

Pemilihan kombinasi titik knot dilakukan dengan melakukan kombinasi dari knot-knot optimum yang telah diperoleh sebelumnya dari hasil penghitungan 1 titik knot, 2 titik knot dan 3 titik knot. Selanjutnya akan dihitung GCV minimum berdasarkan kombinasi yang diperoleh dan dipilih model dengan GCV minimum diantara kombinasi yang ada. Model regresi semiparametrik spline *truncated* dengan menggunakan kombinasi titik knot dengan prediktor satu komponen parametrik dan tiga komponen nonparametrik adalah sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{31} (t_{1i} - K_{21})_+^1 + \theta_{41} (t_{1i} - K_{21})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{32} (t_{2i} - K_{22})_+^1 + \theta_{42} (t_{2i} - K_{32})_+^1 + h(z_i) + \varepsilon_i ; i = 1, 2, \dots, 38$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan kombinasi knot disajikan dalam Tabel 4.6.

Berdasarkan Tabel 4.6 nilai GCV minimum yang dihasilkan adalah sebesar 0,003964 dengan kombinasi titik knot, 2,3 dan lokasi titik knot pada variabel (t_1) yaitu 68,4 (K_{11}) 80,8 (K_{12}) dan untuk (t_2) yaitu 72,1 (K_{21}) 76,074 (K_{22}) 78,062 (K_{23}) sedangkan bandwidth yang diberikan sebesar $\alpha = 0,0427$.

Tabel 4.6 Perbandingan Nilai GCV menggunakan Kombinasi Titik Knot

| SplineKnot | | Kernel | | GCV | Kombinasi | R^2 |
|----------------|----------------|-----------|--|----------|-----------|-------|
| Knot | | Bandwidth | | | | |
| $t_1 = K_{11}$ | $t_2 = K_{21}$ | α | | | | |
| $t_1 = K_{12}$ | $t_2 = K_{22}$ | | | | | |
| $t_1 = K_{13}$ | $t_2 = K_{23}$ | | | | | |
| 68,4 | 72,1 | | | | | |
| 80,8 | 76,074 | 0,0427 | | 0,003964 | 2,3 | 97,04 |
| | 78,062 | | | | | |
| 56,3 | 72,1 | | | | | |
| | 76,074 | 0,04275 | | 0,004046 | 1,3 | 96,98 |
| | 78,062 | | | | | |
| 68,4 | 71,72 | | | | | |
| 80,8 | | 0,04848 | | 0,004088 | 2,1 | 95,12 |
| 57,69 | 71,72 | | | | | |
| 72,345 | | 0,04865 | | 0,004249 | 3,1 | 95,11 |
| 79,673 | | | | | | |

4.2.4 Pemilihan Model Terbaik

Berdasarkan nilai GCV untuk setiap titik knot yang telah dihitung sebelumnya, berikutnya dilakukan pemilihan model terbaik dengan membandingkan nilai GCV yang dihasilkan oleh masing-masing model yang ditunjukkan pada Tabel 4.7.

Tabel 4.7 Nilai GCV Minimum Pada Tiap Model

| Banyaknya Knot | GCV | R-Square |
|----------------------|----------|----------|
| 1 Titik Knot | 0,004164 | 95,06 % |
| 2 Titik Knot | 0,004185 | 96,32 % |
| 3 Titik knot | 0,004115 | 97,05 % |
| Kombinasi Titik Knot | 0,003964 | 97,04 %* |

Tabel 4.7 memperlihatkan bahwa nilai GCV minimum terdapat pada Kombinasi titik knot. Model yang dipilih adalah model Spline dengan tiga titik knot. Setelah diperoleh skor GCV minimum untuk model Spline *truncated* linier, langkah selanjutnya menghitung estimasi untuk model Spline *truncated* linier. Estimasi model Spline *truncated* linier dengan kombinasi titik knot sebagai berikut.

$$\hat{y}_i = -0,17519302433 - 0,00012335311x_{1i} - 0,00079110874t_{1i} + 0,00166672135(t_{1i} - 68,4)_+^1 + 0,00055557378(t_{1i} - 80,80)_+^1 - 0,00329040416t_{2i} - 0,01911809653(t_{2i} - 72,1)_+^1 - 0,08784561457(t_{2i} - 76,07)_+^1 + 0,19710550155(t_{2i} - 78,062)_+^1 + \sum_{i=1}^n \left(\frac{\frac{1}{0,0427} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{z_i - z_{1i}}{0,0427} \right)^2}}{\sum_{i=1}^n \frac{1}{0,0427} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{z_i - z_{1i}}{0,0427} \right)^2}} \right) y_i$$

Model regresi Spline *truncated* linier kombinasi titik knot ini memiliki R² sebesar 97,04%. Hal ini berarti bahwa model ini dapat menjelaskan TFR sebesar 97,04%.

4.2.5 Pengujian Asumsi Residual

Residual pada regresi semiparametrik campuran spline *truncated* dan kernel harus memenuhi identik, independen dan berdistribusi normal (IIDN). Pemeriksaan asumsi residual ini dilakukan untuk mengetahui apakah residual yang terbentuk telah memenuhi asumsi IIDN. Pengujian asumsi residual identik menggunakan uji *glejser* dan pengujian asumsi normal menggunakan *Kolmogorov-Smirnov*.

4.2.5.1 Uji Identik

Identik yang dimaksud disini adalah varians residualnya harus homogen (tidak terdapat heteroskedastisitas). Untuk mendeteksi adanya heteroskedastisitas dilakukan dengan melakukan uji Glejser. Hipotesis yang digunakan untuk uji Glejser adalah sebagai berikut.

$$H_0 : \sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_{38} = \sigma^2$$

$$H_1 : \text{Minimal ada satu } \sigma^2_i \neq \sigma^2, i = 1, 2, \dots, 38$$

Tabel 4.8 ANOVA Uji Glejser

| Sumber Variasi | df | SS | MS | F Hitung | P-Value |
|----------------|----|-----------|------------|-------------|--------------|
| Regresi | 8 | 1,5152299 | 0,16835888 | | |
| Error | 29 | 163,1512 | 5,8268286 | 0,028893742 | 0,9999990294 |
| Total | 37 | 164,66643 | | | |

Tabel 4.8 menunjukkan Nilai Mean Square Regresi (MSR) dan Mean Square Error (MSE) yang dihasilkan pada uji Glejser yakni masing-masing sebesar 0,16835888 dan 5,8268286. Karena p-value (0,9999990294) > nilai alpha (0,05), maka dapat disimpulkan bahwa varians residual memenuhi asumsi identik.

4.2.5.2 Uji Independen

Pengujian asumsi independen dimaksudkan untuk mendeteksi korelasi antar residual. Pemeriksaan dilakukan dengan menggunakan uji *Durbin-Watson*. Hipotesis yang digunakan adalah sebagai berikut:

$H_0: \rho = 0$ (tidak terdapat *autokorelasi*)

$H_1: \rho \neq 0$ (Terdapat *autokorelasi*)

Hasil penghitungan diperoleh $d_{hitung} = 2,004$ dengan $d_{U,0,05/2} = 1,722$ maka diperoleh keputusan gagal tolak H_0 . Hal ini memberikan kesimpulan bahwa antar residual tidak terjadi korelasi. Dengan demikian asumsi independen pada model regresi semiparametrik campuran spline *truncated* dan *kernel* terpenuhi.

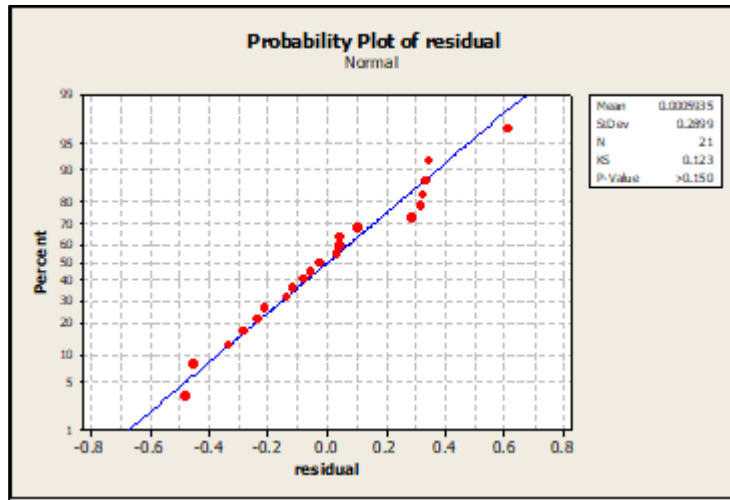
4.2.5.3 Uji Distribusi Normal

Residual dari model regresi semiparametrik Spline *truncated* linier yang diperoleh harus memenuhi asumsi distribusi normal. Pengujian distribusi normal dilakukan dengan uji *Kolmogorov-Smirnov*. Adapun hipotesis yang digunakan dalam pengujian distribusi normal untuk residual ini 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)

Gambar 4.5 menunjukkan bahwa sebaran titik mengikuti garis distribusi normal, artinya distribusi residual mengikuti distribusi normal. Selain itu nilai p-value ($>0,150$) lebih besar dari $(0,05)$ dan nilai $|D| = 0,08979$ lebih kecil dari $q_{(0,05)} = 0,210$ maka residual model memenuhi asumsi distribusi normal.



Gambar 4.5 Hasil Uji Kolmogorov-Smirnov

4.2.6 Pengujian Signifikansi parameter secara Simultan

Pengujian hipotesis untuk menguji signifikansi parameter secara simultan menggunakan hipotesis sebagai berikut:

$$H_0 : \mathbf{C}^T \tilde{\gamma} = \boldsymbol{\tau}$$

$$H_1 : \mathbf{C}^T \tilde{\gamma} \neq \boldsymbol{\tau} .$$

Tabel 4.9 ANOVA Hasil Regresi Semiparametrik campuran spline *truncated* dan *kernel*

| Sumber Variasi | df | SS | MS | F Hitung | P-Value |
|----------------|----|-------------|--------------|-----------|---------|
| Regresi | 8 | 1,5152299 | 0,16835888 | | |
| Error | 29 | 0,046281417 | 0,0016529077 | 101,85619 | 0,000 |
| Total | 37 | 1,5615114 | | | |

Hasil ANOVA pada pengujian hipotesis secara simultan tersaji pada Tabel 4.9. Berdasarkan Tabel 4.9 terlihat bahwa nilai F_{hitung} sebesar 101,85 lebih besar dari nilai $F_{(0,05,8,29)}$ yaitu 2,28 dan nilai p-value sebesar 0,00 lebih kecil dari nilai α (0,05), sehingga diperoleh keputusan tolak H_0 . Maka dapat disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang signifikan dalam model. Model regresi semiparametrik campuran spline *truncated* dengan kombinasi titik knot ini memiliki koefisien determinasi (R^2) sebesar 97,04 persen. Hal ini menunjukkan bahwa variasi dari variabel respon dapat dijelaskan oleh kelima variabel prediktor sebesar 97,04 persen, sedangkan sisanya 02,96 persen dijelaskan oleh faktor lain.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut.

1. Pengujian hipotesis simultan parameter komponen parametrik dan komponen nonparametrik dalam model regresi semiparametrik campuran spline *truncated* dan *kernel* dapat dilakukan dengan menggunakan *Likelihood Ratio Test* dengan formulasi hipotesis sebagai berikut:

$$H_0 : \mathbf{C}^T \tilde{\gamma} = \boldsymbol{\tau}$$

$$H_1 : \mathbf{C}^T \tilde{\gamma} \neq \boldsymbol{\tau}$$

- a. Statistik test untuk pengujian hipotesis:

$$F = \frac{\mathbf{Q}_1 / s}{\mathbf{Q}_2 / (1 + p + (m + r)q)}$$

- b. Distribusi dari statistik uji yang diperoleh mengikuti distribusi-F dengan

$$F \sim F_{(s; 1 + p + (m + r)q)}$$

- c. Daerah penolakan untuk hipotesis H_0 :

$$\mathbf{F} > \mathbf{F}_{(\alpha; s; 1 + (p + (m + r)q))}$$

2. Aplikasi pada data TFR di Provinsi Jawa Timur Tahun 2015 mendapatkan hasil sebagai berikut.
 - a. Model terbaik yang diperoleh adalah dengan menggunakan kombinasi titik knot.
 - b. Dari keempat variabel prediktor yang dianalisis terdapat minimal satu parameter yang berpengaruh terhadap variabel respon.
 - c. Koefisien determinasi (R^2) yang diperoleh adalah sebesar 97,04 persen, sehingga model tersebut sudah layak untuk digunakan.

5.2 Saran

Berdasarkan hasil dan pembahasan estimasi parameter model dalam regresi semiparametrik Campuran Spline *truncated* dan *kernel*, saran yang diberikan adalah:

1. Penelitian ini terbatas pada metode Semiparametrik Campuran Spline *truncated* dan *kernel* untuk pengujian hipotesis simultan. Untuk penelitian lanjut disarankan untuk menggunakan dilakukan pengujian hipotesis secara parsial.
2. Mengembangkan regresi semiparametrik campuran spline *truncated* dan *kernel* dengan spasial, untuk mengetahui pengaruh perbedaan antar wilayah.

DAFTAR PUSTAKA

- Afifah, N., (2017), *Estimator Campuran Kernel dan Deret Fourier Dalam Regresi Nonparametrik (Studi Kasus : Pemodelan Pesentase Penduduk Miskin di Indonesia)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Antoniadis, A., Gregore, G., dan Mckeague, W., Wavelet. (1994), “*Methods for Curve Estimation*”, JASA, Vol.89, No.428, 1994, 1340-1353.
- Antoniadis A., (2007), “*Wavelet methods in statistics: Some recent developments and their applications*”, Statistics Survey, Vol. 1, 16–55.
- BKKBN. 2017., *Profil Kependudukan Jawa Timur Tahun 2017*. Jawa Timur: BKKBN
- BKKBN. 2017., *Buku Saku Parameter Kependudukan Jawa Timur Tahun 2017*. Jawa Timur: BKKBN
- Budiantara, I.N., (2004), *Spline : Historis, Motivasi dan Perannya Dalam Regresi Nonparametrik, Makalah Pembicara Utama pada konferensi Nasional Matematika XII*, Jurusan Matematika, FMIPA, Universitas Udayana, Denpasar.
- Budiantara, I.N., (2005), “*Regresi Spline Linier*”, *Makalah Pembicara Utama pada Seminar Nasional Matematika*, FMIPA Universitas Diponegoro, Semarang.
- Budiantara, I.N., (2009), “*Spline Dalam Regresi Nonparametrik dan Semiparametrik : Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*”, *Pidato Pengukuhan untuk Jabatan Guru Besar*, Institut Teknologi Sepuluh Nopember, ITS Press, Surabaya.
- Budiantara, I.N., Ratnasari, V., Ratna, M and Zain, I. (2015), *The Combination of Spline and Kernel Estimator for Nonparametric Regression and Its Properties*, *Applied Mathematical Science*, 9(122) : 6083-6094.
- Drapper, N. R., dan Smith, H. (1992), *Analisis Regresi Terapan*, PT. Gramedia Pustaka Utama Jakarta.
- Eubank, R., (1999), *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.
- Fendy, A., (2017), *Estimasi Interval Parameter Komponen Parametrik dan Spline Dalam model Regresi Semiparametrik Campuran Spline dan Kernel (Aplikasi : Tingkat Pengangguran Terbuka di Indonesia)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Gujarati, D. N., dan Porter, D. C. (2015). *Dasar-Dasar Ekonometrika Volume 1, 2nd Edition*. Jakarta: Penerbit Salemba Empat.
- Hardle, W., (1990), “*Applied Nonparametric Regression*”, *Cambridge University Press*, New York.
- Hardle, W., (1994). *Applied Nonparametric Regression*. Cambridge University Press. New York.

- Hesikumalasari., (2016), *Model Regresi Semiparametrik Campuran Spline Truncated dan Deret Fourier (Studi Kasus : Angka Harapan Hidup Provinsi Jawa Timur)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Khusniawati., (2017), *Pengujian Hipotesis Parameter Komponen Spline dalam Model Regresi Nonparametrik Campuran Spline dan Kernel* Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Nisa, K., (2017), *Estimasi Interval Parameter Model Spline Truncated Dalam Regresi Semiparametrik (Aplikasi : Pada Data Indeks Pembangunan Indonesia)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Purnomo., A., A., S., I. (2016), *Estimator Campuran Kernel dan Regresi Spline Truncated Linier Multivariabel dalam Regresi Nonparametrik*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Rismal., (2016), *Estimasi Model Campuran Spline Truncated dan Kernel dalam Regresi Nonparametrik Multivariabel*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Rory., (2016), *Regresi Campuran Nonparametrik Spline Linier Truncated dan Fungsi Kernel untuk Pemodelan Data Kemiskinan di Provinsi Papua*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Ruppert., D., Wand, M. P., dan Carrol, R. J., 2003, *Semiparametric Regression*, Cambridge University, United Kingdom.
- Sudiarsa., I. W., Budiantara, I. N., Suhartono, dan Purnami, S. W. 2015. *Combined Estimator Fourier Series and Spline Truncated in Multivariabel Nonparametrik Regression*, Applied Mathematical Science, Vol.9, 2015, no. 100, 4997-5010, HIKARI Ltd.
- Tupen., S.N. (2011), *Uji Hipotesis Dalam Regresi Nonparametrik Spline*, Tesis, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember, Surabaya.
- Wahba,G., (1990) *Spline Models for Observational Data*. Philadelphia:Society
- Wang, Y., 1998, "Spline Smoothing Models with Correlated Error", *Journal of the Royal Statistical Sociaty, Series B*, 5r0, 341-348.
- Zaki, A., (2007), *Pengujian Hipotesis dalam Model Spline pada Regresi Nonparametrik*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.

LAMPIRAN

Lampiran 1. Data yang Digunakan

| No | Y | X1 (UN) | X2 (ASFR) | X3 (IPM) | X4 (AKB) |
|----|------|---------|-----------|----------|----------|
| 1 | 2.01 | 14.63 | 63.00 | 64.92 | 21.21 |
| 2 | 2.13 | 20.50 | 23.90 | 68.16 | 23.89 |
| 3 | 1.94 | 15.65 | 51.00 | 67.25 | 19.66 |
| 4 | 2.34 | 25.88 | 47.50 | 70.07 | 20.35 |
| 5 | 2.42 | 19.77 | 16.20 | 68.13 | 22.23 |
| 6 | 2.40 | 16.46 | 44.30 | 68.91 | 24.75 |
| 7 | 2.31 | 11.37 | 53.00 | 66.63 | 27.81 |
| 8 | 1.85 | 12.50 | 35.30 | 63.02 | 35.13 |
| 9 | 2.05 | 11.60 | 46.80 | 63.04 | 54.01 |
| 10 | 2.22 | 15.05 | 35.70 | 68.08 | 29.07 |
| 11 | 2.09 | 6.87 | 87.00 | 63.95 | 45.59 |
| 12 | 1.78 | 13.94 | 40.30 | 64.53 | 52.30 |
| 13 | 2.04 | 10.77 | 63.20 | 63.83 | 60.51 |
| 14 | 1.90 | 15.55 | 35.70 | 65.04 | 47.47 |
| 15 | 2.00 | 13.06 | 16.30 | 77.43 | 22.19 |
| 15 | 2.20 | 14.25 | 44.90 | 70.85 | 21.64 |
| 17 | 2.14 | 11.91 | 68.20 | 69.59 | 26.56 |
| 18 | 2.28 | 13.16 | 16.00 | 69.90 | 29.30 |
| 19 | 2.03 | 20.01 | 25.80 | 69.39 | 29.75 |
| 20 | 2.22 | 12.52 | 43.00 | 71.39 | 21.26 |
| 21 | 2.10 | 13.26 | 46.70 | 68.32 | 23.79 |
| 22 | 1.98 | 9.80 | 63.40 | 66.17 | 37.50 |
| 23 | 1.82 | 12.90 | 51.80 | 65.52 | 30.31 |
| 24 | 2.04 | 10.94 | 14.70 | 69.84 | 32.39 |
| 25 | 2.12 | 15.41 | 24.80 | 73.57 | 21.26 |
| 26 | 1.89 | 31.77 | 21.70 | 61.49 | 52.56 |
| 27 | 2.45 | 23.11 | 34.70 | 58.18 | 47.28 |
| 28 | 2.03 | 16.5 | 37.20 | 63.10 | 45.97 |
| 29 | 1.52 | 22.95 | 33.80 | 62.38 | 46.06 |
| 30 | 1.94 | 18.46 | 8.20 | 75.67 | 20.86 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 36 | 2.06 | 22.27 | 13.50 | 79.48 | 21.59 |
| 37 | 1.72 | 21.50 | 23.60 | 79.47 | 21.34 |
| 38 | 1.96 | 20.32 | 33.70 | 72.62 | 26.26 |

Lampiran 2. Syntax R Model Semiparametrik Campuran Spline *kernel*

```
function(data,tknot,mx,mxker)
{
  #tknot=number knot point
  #mx=number of parametric
  #mxker=number of non parametric with kernel
  library(pracma)
  alpha=0.05
  n=nrow(data)
  m=ncol(data)
  mnonpar=m-mx-mxker-1
  mk=mnonpar*tknot
  mker=mxker
  y=data[,1];
  x=as.matrix(cbind(1,data[,2:(mx+1)]))
  xnon=as.matrix(data[,((mx+2):(m-mxker))]);
  xker=as.matrix(data[, (m-mxker+1):m])
  if (tknot==1) nk=22 else if (tknot==2) nk=14 else nk=12
  m1.nn = matrix(1,nrow=n,ncol=n)
  satu = matrix(1,nrow=n,ncol=1)
  mi.nn = diag(1,nrow=n,ncol=n)
  knot=matrix(ncol=(mnonpar*tknot),nrow=nk)
  In=diag(n)
  for (i in 1:mnonpar)
    knot[,i]= seq(min(xnon[,i]),max(xnon[,i]),length.out=nk)
  knot=as.matrix(knot[2:(nk-1),])
  nk=nrow(knot)
  knot1=knot

  if (tknot==1){ # build knot point
    knot1=knot1;
  }else if (tknot==2){
    z=(nk*(nk-1)/2);
    knot2=matrix(nrow=(z+1),ncol=mk)
    for (i in 1:mnonpar)
      {
        knot=matrix(nrow=1,ncol=2)
        for (j in 1:(nk-1))
          {
            for ( k in (j+1):(nk))
              {
                xx=cbind(knot1[j,i], knot1[k,i])
                knot=rbind(knot,xx)
              }
            }
          }
    knot2[, (i*2-1):(i*2)]=knot;
```



```

}
knot2=knot2[2:z+1,]
knot1=knot2
}
else {
z=(nk*(nk-1)*(nk-2)/6)
knot2=matrix(nrow=(z+1),ncol=mk)
for (i in 1:mnonpar)
{
knot=matrix(nrow=1,ncol=3)
for (j in 1:(nk-2))
{
for ( k in (j+1):(nk-1))
{
for (g in (k+1):nk)
{
xx=cbind(knot1[j,i] ,knot1[k,i] ,knot1[g,i])
knot=rbind(knot, xx)
}
}
}
knot2[(i*3-2):(i*3)]=knot
}
knot2=knot2[2:z+1,]
knot1=knot2;
}
knot=knot1
nk=nrow(knot)
res=matrix(ncol=nk,nrow=n)
yhat1=matrix(ncol=nk,nrow=n)
matGCV=matrix(ncol=1,nrow=nk)
matMSE=matrix(ncol=1,nrow=nk)
matRsqr=matrix(ncol=1,nrow=nk)
matalfa=matrix(ncol=mxker,nrow=nk)
beta=matrix(ncol=(1+mx),nrow=nk)
ba_beta=matrix(ncol=(1+mx),nrow=nk)
bb_beta=matrix(ncol=(1+mx),nrow=nk)
teta=matrix(ncol=(mnonpar+(mnonpar*tknot)),nrow=nk)
ba_teta=matrix(ncol=(mnonpar+(mnonpar*tknot)),nrow=nk)
bb_teta=matrix(ncol=(mnonpar+(mnonpar*tknot)),nrow=nk)
nbatateta=(1+mx)+(mnonpar+(mnonpar*tknot))
Ix=diag((1+mx))
Ixnon=diag((mnonpar+(mnonpar*tknot)))
for(i in 1:nk)
{
cat('loading... ',i/nk*100,"% ", "\n")
alfa=c(rep(0.05,mker))

```

```

matknot=matrix(ncol=(mnonpar*tknot),nrow=n)
GCVmin = function(alfa)
{
sumWZ = 0
for(j in 1:mker)
{
xkerdiag = diag(xker[,j])
V = m1.nn%*%xkerdiag
Z = (t(V)-V)/alfa[j]
K = 1/sqrt(2*pi)*exp(-1/2*Z^2) #fungsi gaussian
KA = (1/alfa[j])*K
Wbawah = diag(c(1/n*KA%*%satu))%*%m1.nn
WZ = 1/n*KA/Wbawah
sumWZ = sumWZ+WZ
}
Xker = sumWZ/mker

for ( k in 1:(tknot*mnonpar))
{
if (tknot==1) { b=k;
} else if (tknot==2) { if (mod(k,2)==1) b=floor(k/2)+1 else b=k/2
} else {b=ceil(k/3);}
matknot[,k]=pmax(0,(xnon[,b]-knot[i,k]));
}
Xnon=as.matrix(cbind(xnon,matknot))

invXnon=pinv(t(Xnon)%*%Xnon)%*%t(Xnon)
invx=pinv(t(x)%*%x)%*%t(x)

B1=pinv(Ix-(invx%*%Xnon%*%invXnon%*%x))
B2=invx%*%(In-Xker)-invx%*%Xnon%*%invXnon%*%(In-Xker)
Beta1=B1%*%B2

T1=pinv(Ixnon-(invXnon%*%x%*%invx%*%Xnon))
T2=invXnon%*%(In-Xker)-invXnon%*%x%*%invx%*%(In-Xker)
Teta1=T1%*%T2
yhat = (x%*%Beta1+Xnon%*%Teta1+Xker)%*%y
SSE = sum((y-yhat)^2)
MSE = SSE/n
GCV = MSE/((n^1*sum(diag(mi.nn-
(x%*%Beta1+Xnon%*%Teta1+Xker)))^2))
return(GCV)
}
optimasi = suppressWarnings(nlminb(c(rep(0.05,mker)),GCVmin,lower=0))
matalfa[i,] = optimasi$par
matGCV[i] = optimasi$objective
alfa = matalfa[i,]

```

```

sumWZ = 0
for(j in 1:mker)
{
xkerdiag = diag(xker[,j])
V = m1.nn%%xkerdiag
Z = (t(V)-V)/alfa[j]
K = 1/sqrt(2*pi)*exp(-1/2*Z^2) #fungsi gaussian
KA = (1/alfa[j])*K
Wbawah = diag(c(1/n*KA%%satu))%%m1.nn
WZ = 1/n*KA/Wbawah
sumWZ = sumWZ+WZ
}
Xker = sumWZ/mker
for ( k in 1:(tknot*mnonpar))
{
if (tknot==1) { b=k;
} else if (tknot==2) { if (mod(k,2)==1) b=floor(k/2)+1 else b=k/2
} else {b=ceil(k/3);}
matknot[,k]=pmax(0,(xnon[,b]-knot[i,k]));
}
Xnon=as.matrix(cbind(xnon,matknot))
invXnon=pinv(t(Xnon)%Xnon)%t(Xnon)
invx=pinv(t(x)%x)%t(x)

B1=pinv(Ix-(invx%%Xnon%invXnon%x))
B2=invx%(In-Xker)-invx%Xnon%invXnon%(In-Xker)
Beta1=B1%B2

T1=pinv(Ixnon-(invXnon%x%invx%Xnon))
T2=invXnon%(In-Xker)-invXnon%x%invx%(In-Xker)
Teta1=T1%T2
yhat1[,i] = (x%%Beta1+Xnon%%Teta1+Xker)%y
res[,i]=yhat1[,i]-y
beta[,i]=Beta1%y
teta[,i]=Teta1%y
SSE = sum((y-yhat1[,i])^2)
SSR=sum((yhat1[,i]-mean(y))^2)
SST=SSR+SSE
matRsq[i]=(SSR/(SST))*100
matMSE[i] = SSE/n

}

```

```
dataprint = data.frame(cbind(matGCV,matRsqr,matMSE,beta,teta,knot,matalfa))  
  
hasil=list(Matriks.GCV=matGCV,Rsquard=matRsqr,MSE=matMSE,nilaibeta=b  
eta,  
nilaiteta=teta,titikknot=knot,malfa=matalfa)  
return(hasil)  
}
```

BIOGRAFI PENULIS



Arip Ramadan lahir di Ciamis pada tanggal 12 Maret 1993. Penulis merupakan anak dari pasangan bapak Anan dan ibu Enung Runansih Sriwulandari. Anak kedua dari tiga bersaudara, anak pertama bernama Irfan Faturochman dan anak ketiga bernama Ganjar Ahmad Rifa'i. Penulis memulai jenjang pendidikan formal di Sekolah Dasar Negeri 2 Cileungsir, Kabupaten Ciamis pada tahun 1999-2005. Kemudian penulis melanjutkan pendidikan ke Sekolah Menengah Pertama Negeri 2 Rancah tahun 2005-2008. Penulis menyelesaikan pendidikan yang selanjutnya di Sekolah Menengah Atas Negeri 3 Ciamis pada tahun 2008- 2011. Penulis melanjutkan pendidikan Sarjana (S1) di Institut Sains & Teknologi Amanat Keluarga Pejuang Republik Indonesia (IST AKPRIND) Yogyakarta pada program studi Statistika sampai selesai pada tahun 2016. Pada tahun yang sama melanjutkan studi Pascasarjana (S2) di Institut Teknologi Sepuluh Nopember (ITS) Suraabaya, Fakultas Matematika, Komputasi, dan Sains Data (FMKSD), Departemen Statistika. Pembaca yang ingin memberikan kritik, saran dan pertanyaan mengenai penelitian ini bisa menghubungi penulis melalui email: aripramadan@gmail.com.