



TESIS - SS14 2501

**ESTIMASI INTERVAL PARAMETER KOMPONEN  
PARAMETRIK DAN SPLINE DALAM MODEL  
REGRESI SEMIPARAMETRIK CAMPURAN  
SPLINE DAN KERNEL**

(Aplikasi : Tingkat Pengangguran Terbuka di Indonesia)

FENDY ARDYANTO  
NRP. 06211650017012

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

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



TESIS - SS14 2501

**ESTIMASI INTERVAL PARAMETER KOMPONEN  
PARAMETRIK DAN SPLINE DALAM MODEL  
REGRESI SEMIPARAMETRIK CAMPURAN  
SPLINE DAN KERNEL**

(Aplikasi : Tingkat Pengangguran Terbuka di Indonesia)

FENDY ARDYANTO  
NRP. 06211650017012

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

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



TESIS - SS14 2501

**ESTIMATION INTERVAL OF PARAMETER  
PARAMETRIC AND SPLINE COMPONENTS  
IN SEMIPARAMETRIC REGRESSION  
MIXED SPLINE DAN KERNEL**

(Application : Open Unemployment Rate in Indonesia)

FENDY ARDYANTO  
NRP. 06211650017012

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

MAGISTER PROGRAM  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTATION, AND DATA SCIENCE  
INSTITUTE OF TECHNOLOGY SEPULUH NOPEMBER  
SURABAYA  
2018

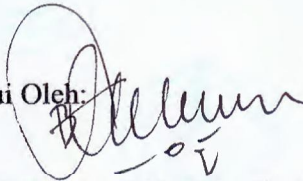
**ESTIMASI INTERVAL PARAMETER KOMPONEN  
PARAMETRIK DAN SPLINE DALAM MODEL REGRESI  
SEMIPARAMETRIK CAMPURAN SPLINE DAN KERNEL  
(Aplikasi : Tingkat Pengangguran Terbuka di Indonesia)**

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Sains (M.Si)  
di  
Institut Teknologi Sepuluh Nopember  
Oleh:

**FENDY ARDYANTO  
NRP. 06211650017012**

Tanggal Ujian : 19 Januari 2018  
Periode Wisuda : Maret 2018


Disetujui Oleh:

- 
1. Prof. Dr. Drs. I Nyoman Budiantara, M.Si. (Pembimbing I)  
NIP. 19650603 198903 1 003
  2. Dr. Dra. Ismaini Zain, M.Si. (Pembimbing II)  
NIP. 19600525 198803 2 001
  3. Dr. Vita Ratnasari, S.Si., M.Si. (Penguji)  
NIP. 19700910 199702 2 001
  4. Dr. Margaretha Ari Anggorowati, S.Kom., M.T. (Penguji)  
NIP. 19720222 199803 2 002



Dekan

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

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

**ESTIMASI INTERVAL PARAMETER  
KOMPONEN PARAMETRIK DAN SPLINE  
DALAM MODEL REGRESI SEMIPARAMETRIK  
CAMPURAN SPLINE DAN KERNEL  
(Aplikasi : Tingkat Pengangguran Terbuka di Indonesia)**

Nama Mahasiswa : Fendy Ardyanto  
NRP : 06211650017012  
Pembimbing : Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
Co-Pembimbing : Dr. Dra. Ismaini Zain, M.Si

**ABSTRAK**

Estimasi interval merupakan salah satu bagian penting dari inferensi statistik. Estimasi interval untuk parameter dalam regresi semiparametrik dapat digunakan untuk menentukan variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon. Dalam penelitian ini dikembangkan estimasi interval untuk parameter model dengan menggunakan beberapa variabel prediktor yang mempunyai pola hubungan dengan variabel respon yang berbeda-beda. Pendekatan regresi semiparametrik campuran spline *truncated* dan kernel digunakan untuk menjelaskan pola hubungan tersebut. Untuk menyelesaikan estimasi interval terpendek parameter model, digunakan pendekatan *pivotal quantity*. Dari hasil penelitian ini didapatkan rumusan estimasi interval untuk parameter saat varians diketahui maupun saat varians tidak diketahui.

Kajian estimasi interval untuk parameter model regresi semiparametrik campuran spline *truncated* dan kernel diaplikasikan pada data Tingkat Pengangguran Terbuka di Indonesia pada Tahun 2016. Metode pemilihan titik knot dan bandwidth optimum menggunakan metode Generalized Cross Validation (GCV). Model terbaik yang terbentuk adalah model menggunakan satu titik knot dengan GCV minimum 0,04054 dan koefisien determinasi disesuaikan ( $R^2$  adj) sebesar 92,53 persen. Variabel yang berpengaruh secara signifikan adalah tingkat partisipasi angkatan kerja pada komponen parametrik, persentase penduduk 15 tahun keatas lulus SMA/SMK, kepadatan penduduk dan laju pertumbuhan ekonomi pada komponen spline. Dari hasil penelitian ini diharapkan dapat dijadikan masukan bagi pemerintah untuk mengambil kebijakan dalam rangka menurunkan Tingkat Pengangguran Terbuka secara menyeluruh.

**Kata kunci** : analisis regresi, tingkat pengangguran terbuka, *pivotal quantity*, estimasi interval, semiparametrik, spline *truncated*, kernel

**ESTIMATION INTERVAL OF PARAMETER  
PARAMETRIC AND SPLINE COMPONENTS  
IN SEMIPARAMETRIC REGRESSION  
MIXED SPLINE AND KERNEL  
(Application: Open Unemployment Rate in Indonesia)**

Name : Fendy Ardyanto  
NRP : 06211650017012  
Supervisor : Prof. Dr. Drs. I Nyoman Budiantara, M.Si  
Co-Supervisor : Dr. Dra. Ismaini Zain, M.Si

**ABSTRACT**

Interval estimates are an important part of statistical inference. Interval estimates for parameters in semiparametric regression can be used to determine predictor variables that significantly influence response variables. In this research, interval estimation is developed for model parameters by using several predictor variables that have relationship with response variable in different pattern. A mixture model of spline truncated and kernel in semiparametric regression approach is used to explain the pattern of the relationship. To complete the shortest interval estimate of model parameters, the pivotal quantity approach is used. From the results of this study obtained the formulation of interval estimation for parameters when the variance is known and when the variance is unknown.

Interval estimation studies for semiparametric mixed regression spline and kernel semiparametric regression parameters were applied to the Unemployment Rate data in Indonesia in 2016. The method of selecting knots and optimum bandwidth using Generalized Cross Validation (GCV) method. The best model is model using one point knot with minimum GCV 0,04054 and coefficient of determination adjusted ( $R^2$  adj) equal to 92,53 percent. Significant variables that influence significantly are labor force participation rate on parametric component, the percentage of population 15 years and above graduated from SMA / SMK, population density and economic growth rate on spline component.. From the results of this study is expected to be an input for the Government to take policy in order to reduce the Unemployment Rate Open as a whole.

**Keywords:** regression analysis, open unemployment rate, pivotal quantity, interval estimate, semiparametric regression, spline truncated, kernel

## KATA PENGANTAR

Segala puji milik Allah SWT, Dzat Yang Maha Esa, yang telah memberikan karunia nikmat kepada penulis sehingga tesis yang berjudul berjudul *“Estimasi Interval Parameter Komponen Parametrik dan Spline dalam Model Regresi Semiparametrik Campuran Spline dan Kernel”* dengan baik dan tepat waktu.

Keberhasilan penyusunan tesis ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Untuk itu, pada kesempatan ini teriring rasa syukur dan doa, penulis menyampaikan ucapan terima kasih kepada :

1. Badan Pusat Statistik (BPS) yang telah memberi kesempatan serta beasiswa kepada penulis untuk melanjutkan studi program S2 di ITS.
2. Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si. dan Ibu Dr. Dra. Ismaini Zain, M.Si. selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, masukan, saran dan motivasi dengan penuh kesabaran selama penyusunan tesis ini.
3. Ibu Dr. Vita Ratnasari, S.Si., M.Si., dan Ibu Ibu Dr. Margaretha Ari Anggorowati, S.Kom, MT. selaku penguji yang telah banyak memberikan saran dan masukan untuk menjadikan tesis ini menjadi lebih baik.
4. Bapak Dr. Suhartono, M.Sc., selaku Ketua Jurusan Statistika dan Bapak Dr. rer.pol. Heri Kuswanto, M.Si., selaku Ketua Program Studi Pascasarjana Jurusan Statistika FMKSD ITS atas arahan dan bantuannya selama penulis menempuh pendidikan di Program Magister Jurusan Statistika ITS.
5. Ibu Dr. Irhamah, M.Si., selaku dosen wali, seluruh Bapak/Ibu dosen pengajar yang telah memberikan ilmu dan pengalaman yang bermanfaat kepada penulis, serta segenap karyawan dan keluarga besar Jurusan Statistika FMKSD ITS atas segala dukungan dan bantuannya.
6. Teristimewa untuk Istriku tercinta, Middia Martanti Dewi yang selalu sabar dalam mendidik anak-anak, senantiasa mendoakan, mendukung serta memberi semangat pada penulis. Untuk Anak-anakku tersayang Rasya Muhammad

Athaya dan Fadlan Muhammad Hamizan yang telah memberikan warna kehidupan, penyejuk hati dan penyemangat jiwa bagi penulis.

7. Bapak Abdul Kadir dan Ibu Murdiah (orang tuaku tercinta) yang telah membesarkan, mendidik dan senantiasa mendoakan untuk kebaikan anak-anaknya. Ibu Eni Suciati (mertua penulis) yang turut mendukung dan mendoakan untuk kebaikan penulis.
8. Teman-teman BPS Batch-10 : Mas Prpto ,Mas Umam (Leader), Mas Aniq, Mas Taufiq, Mas Rafael, Dek Sony, Mbak Prih, Mbak Ratna (Si Genius), Mba Reny, Mba Sartika, Mbak Imra, Mbak Mike, Dek Ratih. Terima kasih atas segala bantuan, kebersamaan dan kekompakannya selama menjalani pendidikan di ITS.
9. Teman-teman reguler angkatan 2016 Mbak Vita, Dek Zahro (Nonpar Squad), Pak Irul dan Mba Mia (admin pasca), Mba Linda (perpustakaan/RBS) serta semua pihak yang tidak bisa disebutkan satu per satu, Penulis menyampaikan rasa terima kasih atas bantuan dan dukungannya selama menjalani pendidikan di ITS.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna, oleh karena itu, kritik maupun saran yang bersifat membangun sangat penulis harapkan demi perbaikan dalam penulisan di masa yang akan datang. Akhirnya, penulis berharap mudah-mudahan tesis ini bermanfaat untuk semua pihak yang memerlukan.

Surabaya, Januari 2017

Penulis



## DAFTAR ISI

|  | Halaman |
|--|---------|
| HALAMAN JUDUL.....   | i       |
| LEMBAR PENGESAHAN .....  | ii      |
| ABSTRAK .....  | iii     |
| ABSTRACT .....   | v       |
| KATA PENGANTAR .....   | vii     |
| DAFTAR ISI.....  | ix      |
| DAFTAR TABEL.....  | xi      |
| DAFTAR GAMBAR .....  | xiii    |
| DAFTAR LAMPIRAN.....   | xv      |
| BAB 1 PENDAHULUAN .....  | 1       |
| 1.1 Latar Belakang .....   | 1       |
| 1.2 Rumusan Masalah .....  | 5       |
| 1.3 Tujuan Penelitian.....   | 5       |
| 1.4 Manfaat Penelitian.....  | 5       |
| 1.5 Batasan Masalah.....   | 6       |
| BAB 2 TINJAUAN PUSTAKA .....   | 7       |
| 2.1 Regresi Parametrik .....   | 7       |
| 2.2 Regresi Nonparametrik Spline <i>Truncated</i> .....                      | 8       |
| 2.3 Regresi Nonparametrik Kernel.....  | 10      |
| 2.4 Regresi Semiparametrik Campuran Spline <i>Truncated</i> dan Kernel ..... | 12      |
| 2.5 <i>Generalized Cross Validation</i> (GCV).....                           | 13      |
| 2.6 Estimasi Interval.....   | 13      |
| 2.7 <i>Pivotal Quantity</i> .....  | 15      |
| 2.8 Koefisien Determinasi ( $R^2$ ) .....                                    | 15      |
| 2.9 Teorema Dasar Terkait Aljabar Matriks .....                              | 16      |
| 2.10Tingkat Pengangguran Terbuka (TPT).....                                  | 18      |
| 2.11Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka.....        | 20      |
| BAB 3 METODOLOGI PENELITIAN.....   | 25      |
| 3.1 Sumber Data .....  | 25      |
| 3.2 Variabel Penelitian .....  | 25      |
| 3.3 Struktur Data .....  | 25      |
| 3.4 Definisi Operasional.....  | 26      |
| 3.5 Tahapan Penelitian .....   | 27      |

|   |    |
|---|----|
| BAB 4 HASIL DAN PEMBAHASAN .....  | 33 |
| 4.1 Model Regresi Semiparametrik Campuran Spline dan Kernel.....  | 33 |
| 4.1.1 Estimasi Kurva Regresi Semiparametrik Campuran Spline dan<br>Kernel .....   | 36 |
| 4.1.2 Estimasi Interval Untuk Parameter Regresi Semiparametrik<br>Campuran Spline dan Kernel .....                        | 40 |
| 4.1.3 Estimasi Interval untuk $\sigma^2$ Diketahui.....   | 44 |
| 4.1.4 Estimasi Interval untuk $\sigma^2$ Tidak Diketahui .....  | 46 |
| 4.2 Aplikasi Pada Data Tingkat Pengangguran Terbuka di Indonesia .....  | 54 |
| 4.2.1 Analisis Deskriptif Data .....  | 54 |
| 4.2.2 Memodelkan TPT Menggunakan Regresi Semiparametrik Campuran<br>Spline <i>Truncated</i> dan Kernel.....               | 57 |
| 4.2.2.1 Menentukan Variabel Komponen Parametrik dan<br>Nonparametrik.....   | 57 |
| 4.2.2.2 Memodelkan Tingkat Pengangguran Terbuka dengan Satu<br>Titik Knot .....   | 61 |
| 4.2.2.3 Memodelkan Tingkat Pengangguran Terbuka dengan<br>Komponen Spline <i>Truncated</i> Dua Titik Knot .....           | 62 |
| 4.2.2.4 Memodelkan Tingkat Pengangguran Terbuka dengan<br>Komponen Spline <i>Truncated</i> Tiga Titik Knot.....           | 63 |
| 4.2.2.5 Memodelkan Tingkat Pengangguran Terbuka dengan<br>Komponen Spline <i>Truncated</i> Kombinasi Titik Knot.....      | 64 |
| 4.2.2.6 Estimasi Interval Parameter Model Semiparametrik<br>Campuran Spline <i>Truncated</i> dan Kernel .....             | 65 |
| 4.2.2.7 Model Semiparametrik Campuran Spline <i>Truncated</i> dan<br>Kernel dengan Titik Knot dan Bandwidth Optimum ..... | 66 |
| 4.2.3 Pemeriksaan Asumsi Residual .....   | 71 |
| 4.2.3.1 Pemeriksaan Asumsi Identik .....  | 71 |
| 4.2.3.2 Pemeriksaan Asumsi Independen.....  | 71 |
| 4.2.3.3 Pemeriksaan Asumsi Distribusi Normal .....  | 72 |
| 4.2.4 Perbandingan Model .....  | 73 |
| BAB 5 KESIMPULAN DAN SARAN .....  | 77 |
| 5.1 Kesimpulan .....  | 77 |
| 5.2 Saran   | 78 |
| DAFTAR PUSTAKA.....   | 81 |
| LAMPIRAN .....  | 85 |
| BIOGRAFI PENULIS .....  | 93 |

## DAFTAR TABEL

|            |   |    |
|------------|---|----|
| Tabel 3.1  | Variabel Penelitian .....   | 25 |
| Tabel 3.2  | Struktur Data Penelitian .....  | 26 |
| Tabel 4.1  | Statistik Deskriptif Variabel Respon dan Variabel Prediktor   | 55 |
| Tabel 4.2  | GCV dari Model dengan Komponen 1 Parametrik 4 Spline dan 1 Kernel ..  | 58 |
| Tabel 4.3  | GCV dari Model dengan Komponen 1 Parametrik 3 Spline dan 2 Kernel ...   | 59 |
| Tabel 4.4  | GCV dari Model dengan Komponen 1 Parametrik 2 Spline dan 3 Kernel ..  | 59 |
| Tabel 4.5  | GCV dari Model dengan Komponen 1 Parametrik 1 Spline dan 4 Kernel ..  | 60 |
| Tabel 4.6  | Ringkasan Penentuan Komponen Parametrik dan Nonparametrik .....   | 61 |
| Tabel 4.7  | Perbandingan Nilai GCV dengan menggunakan Satu Titik Knot.....  | 61 |
| Tabel 4.8  | Perbandingan Nilai GCV dengan menggunakan Dua Titik Knot .....  | 62 |
| Tabel 4.9  | Perbandingan Nilai GCV dengan menggunakan Tiga Titik Knot.....  | 63 |
| Tabel 4.10 | Perbandingan Nilai GCV dengan menggunakan Kombinasi Titik Knot.....   | 65 |
| Tabel 4.11 | Estimasi Interval untuk Parameter Komponen Parametrik dan Spline dalam Model Semiparametrik Campuran Spline <i>Truncated</i> dan Kernel ..... | 66 |
| Tabel 4.12 | Perbandingan Nilai GCV Minimum Tiap Model.....  | 67 |
| Tabel 4.13 | Hasil Uji Glejser.....  | 71 |
| Tabel 4.14 | Nilai $R^2$ adjusted.....   | 73 |

## DAFTAR GAMBAR

|   |    |
|---|----|
| Gambar 2.1 Bagan Konsep Ketenagakerjaan.....  | 19 |
| Gambar 2.2 Diagram Konseptual Penelitian.....   | 23 |
| Gambar 3.1 Tahapan Analisis untuk Mendapatkan Estimasi Interval untuk<br>Parameter Komponen Parametrik dan Komponen Spline dalam<br>Regresi Semiparametrik Campuran Spline dan Kernel. .... | 30 |
| Gambar 3.2 Tahapan Analisis Estimasi Interval pada Aplikasi data TPT.....   | 32 |
| Gambar 4.1 Peta TPT Menurut Provinsi di Indonesia .....   | 55 |
| Gambar 4.2 Scatter Plot antara TPT dengan Variabel Prediktor .....  | 57 |
| Gambar 4.3 Peta Hubungan TPT terhadap Persentase Penduduk 15+ yang lulus<br>SMA/SMK.....  | 68 |
| Gambar 4.4 Hubungan antara persentase penduduk 15+ yang lulus SMA/SMK<br>dengan TPT.....  | 69 |
| Gambar 4.5 Peta TPT terhadap Laju Pertumbuhan Ekonomi.....  | 70 |
| Gambar 4.6 Hubungan Antara Laju Pertumbuhan Ekonomi terhadap TPT .....  | 70 |
| Gambar 4.7 Plot Autocorrelation Function (ACF) Residual .....   | 72 |
| Gambar 4.8 Probability Plot Residual.....   | 73 |
| Gambar 4.9 Perbandingan Hasil Estimasi beberapa Metode .....  | 74 |

## DAFTAR LAMPIRAN

|  |    |
|--|----|
| Lampiran 1. Data Tingkat Pengangguran Terbuka di Indonesia Tahun 2016 dan Faktor-Faktor yang diduga Berpengaruh..... | 85 |
| Lampiran 2. <i>Scatter Plot</i> Variabel Respon terhadap Variabel Prediktor .....                                    | 87 |
| Lampiran 3. Sintax R Model Semiparametric Campuran Spline Truncated dan Kernel.....                                  | 88 |

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Salah satu metode statistik yang digunakan untuk mengetahui hubungan antara variabel respon dan variabel prediktor adalah analisis regresi. 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). Sementara jika pola hubungan antara variabel respon dan variabel prediktor tidak diketahui bentuknya dan tidak tersedianya informasi terkait pola data, maka disarankan untuk menggunakan pendekatan dengan regresi nonparametrik. Beberapa contoh model regresi nonparametrik yang banyak digunakan antara lain adalah histogram, *Kernel* (Hardle,1994), *Spline* (Wahba,1990), Deret *Fourier* (Antoniadis, 1994), dan *Wavelets* (Antoniadis,2007). Dalam perkembangannya diantara beberapa model regresi nonparametrik tersebut, *Spline* dan *Kernel* lebih banyak diminati dan mendapatkan perhatian dari para peneliti regresi nonparametrik. Jika hubungan antar variabelnya sebagian diketahui dan sebagian lainnya tidak diketahui polanya, maka pendekatan dengan regresi semiparametrik lebih baik digunakan (Budiantara, 2009).

Dalam beberapa penelitian sebenarnya sering dijumpai kasus-kasus dimana masing-masing variabel prediktor mempunyai pola data yang berbeda. Dengan hanya menggunakan satu bentuk estimator dalam mengestimasi kurva regresi, berakibat estimator yang diperoleh kurang sesuai dengan pola data. 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.

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 sedangkan Nisa (2017) mengestimasi model semiparametrik campuran Spline *Truncated* dan Deret Fourier. Dari penelitian keduanya hanya sebatas mencari estimasi titik dari kurva regresi dan belum dilakukan inferensi lebih lanjut terkait estimasi interval.

Dalam teori estimasi, estimasi titik memegang peranan penting dalam pengembangan statistika inferensia, tetapi secara probabilitas kesalahan dari estimasi titik ini sangat besar. Oleh karena itu diperlukan estimasi interval yang memiliki probabilitas kesalahan yang lebih kecil dengan menggunakan batas atas dan batas bawah dengan tingkat kesalahan tertentu. Estimasi interval bisa digunakan untuk kurva regresi maupun parameter yang dihasilkan. Estimasi interval untuk parameter dapat digunakan untuk mengetahui variabel prediktor yang signifikan terhadap variabel respon. Adapun kriteria estimasi interval terbaik adalah yang memiliki interval terpendek dengan tingkat kesalahan tertentu.

Beberapa penelitian mengenai estimasi interval diantaranya dilakukan oleh Wahba (1983) dan Wang (1998) dengan menggunakan pendekatan bayesian yang melibatkan distribusi *prior improper* sehingga secara matematis sulit dilakukan. Pendekatan dengan *pivotal quantity* dinilai lebih mudah dibandingkan pendekatan bayesian dalam hal estimasi interval, dimana pada pendekatan ini tidak melibatkan distribusi *prior* sehingga diperoleh model yang sederhana dan inferensi statistika yang relatif mudah (Eubank, 1988). Dalam beberapa tahun terakhir, pendekatan ini digunakan oleh peneliti untuk melakukan estimasi interval regresi nonparametrik, diantaranya Nafi' dan Budiantara (2008), Intansari (2016), dan Setiawan (2017) sedangkan Loklomin (2017) melakukan estimasi interval untuk parameter model pada regresi semiparametrik.

Namun dari beberapa penelitian tersebut hanya terbatas pada regresi dengan estimator tunggal, baik dalam regresi nonparametrik maupun semiparametrik. Estimasi interval untuk parameter model regresi dengan estimator

campuran yang memiliki model yang lebih kompleks belum banyak dikembangkan. Oleh karena itu dalam penelitian ini akan dikembangkan estimasi interval untuk parameter model semiparametrik campuran spline *truncated* dan kernel dengan pendekatan *pivotal quantity*. Adapun estimasi interval parameter yang dimaksud hanya untuk komponen parametrik dan spline saja. Penerapan estimasi interval untuk parameter pada penelitian ini akan diterapkan pada data Tingkat Pengangguran Terbuka di Indonesia.

Angka Pengangguran merupakan salah satu tolak ukur keberhasilan pembangunan terutama dibidang ketenagakerjaan. Tingginya angka pengangguran diakibatkan karena jumlah angkatan kerja yang ada belum mampu terserap secara optimal, hal ini disebabkan karena terbatasnya lapangan kerja yang ada. Indikator utama yang digunakan untuk mengukur angka pengangguran adalah Tingkat Pengangguran Terbuka (TPT) yang merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja.

Berdasarkan data ILO, pada tahun 2016 Indonesia menduduki peringkat 2 terbesar berdasarkan tingkat pengangguran (*Unemployment Rate*) antar negara ASEAN yaitu sebesar 5,6 persen, hanya Filipina yang memiliki tingkat pengangguran yang lebih besar yaitu sebesar 5,9. Dalam kurun waktu 5 tahun terakhir Tingkat Pengangguran Terbuka di Indonesia menunjukkan kecenderungan menurun tetapi belum konsisten, masih terjadi kenaikan di beberapa tahun terakhir.. Jika dilihat pada tahun 2016, sebaran tingkat pengangguran antar provinsi menunjukkan disparitas yang cukup tinggi, yang terbesar terdapat di Provinsi Banten yaitu 8,89 dan yang terkecil ada di Provinsi Bali yaitu 1,89. Wilayah yang merupakan pusat pertumbuhan ekonomi diharapkan mampu membuka lapangan usaha baru dan menekan tingkat pengangguran. Akan tetapi yang terjadi adalah tingginya tingkat pengangguran di wilayah tersebut, seperti Banten dan Jawa Barat. Terdapat 7 juta lebih penduduk yang berstatus pengangguran dan jumlahnya akan semakin bertambah seiring dengan meningkatnya jumlah angkatan kerja yang tidak diimbangi oleh lapangan kerja yang memadai dan sesuai.

Beberapa penelitian sebelumnya mengenai TPT pernah dilakukan oleh Tantri dan Ratnasari (2016) dan (Astuti dkk, 2017) menggunakan regresi panel yang menghasilkan beberapa variabel seperti Upah Minimum, Jumlah Penduduk,



TPAK, pertumbuhan ekonomi, Angka Partisipasi Kasar SMA-Sederajat, dan Angka Melek Huruf berpengaruh secara signifikan terhadap TPT. Alghofari (2010) dan Rachim (2013) juga menggunakan pendekatan parametrik dalam penelitiannya, yaitu menggunakan regresi linier berganda. (Sari dkk, 2012) menggunakan pendekatan nonparametrik regresi spline multivariabel dalam pemodelan pengangguran terbuka di Jawa Timur. Hubungan antara variabel respon dan variabel prediktor yang digunakan tidak diketahui pola datanya sehingga didekati dengan nonparametrik spline. Rismal (2016) menggunakan dua pendekatan nonparametrik yang berbeda dalam memodelkan TPT di Jawa Barat. Sebagian variabel prediktor yang digunakan didekati dengan fungsi spline karena pola data yang berubah pada sub-sub interval, sedangkan variabel lainnya didekati dengan fungsi kernel dikarenakan tidak mempunyai pola data yang jelas.

Pendekatan menggunakan *scatter plot* dapat dilakukan sebagai studi awal untuk menentukan metode yang akan digunakan. Apabila hasil *scatter plot* menunjukkan kecenderungan data mengikuti suatu pola seperti linier, kudratik, kubik dan seterusnya maka dapat digunakan dengan regresi parametrik. Dalam keadaan tertentu, pola data hasil *scatter plot* bisa jadi tidak jelas atau mungkin terdapat beberapa pola data yang berbeda. Jika terjadi hal demikian maka pendekatan dengan parametrik maupun nonparametrik saja dipandang kurang tepat dan bijaksana. Salah satu analisis regresi yang dapat digunakan dalam kasus ini adalah regresi semiparametrik. Pola data TPT dan faktor-faktor yang diduga mempengaruhinya dapat dilihat pada Lampiran 1.

Hasil *scatter plot* pada Lampiran 1, untuk dugaan awal hubungan antara TPT dan TPAK menunjukkan kecenderungan linier, semakin tinggi TPAK mengakibatkan TPT cenderung mengalami penurunan. Pola hubungan antara TPT dan Persentase lulus SMA-Sederajat menunjukkan perubahan pada sub-sub interval, terlihat pada interval sebelum 40 cenderung turun melambat dan diatas interval 40 cenderung naik. Sedangkan pola hubungan antara TPT dan variabel lainnya terlihat tidak berpola.

Berdasarkan penelitian sebelumnya dan eksplorasi data yang dilakukan, terdapat beberapa pola data dan pendekatan yang digunakan dalam menentukan variabel yang berpengaruh terhadap TPT. Oleh karena itu analisis regresi

semiparametrik campuran spline dan kernel diharapkan mampu mengakomodir pola data yang berbeda sehingga dihasilkan model yang tepat. Sebagai pengembangannya, penelitian ini akan difokuskan dalam pengkajian estimasi interval parameter model untuk mendapatkan interval konfidensi terpendek, sehingga dapat diketahui variabel yang berpengaruh secara signifikan. Adapun variabel prediktor yang digunakan dan diduga berpengaruh adalah Tingkat Partisipasi Angkatan Kerja (TPAK), Upah Minimum Provinsi (UMP), Persentase penduduk usia 15 + yang lulus SMA, Kepadatan Penduduk, Investasi dan Laju Pertumbuhan Penduduk.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan diatas, maka rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana mendapatkan estimasi interval terpendek parameter komponen parametrik dan spline dalam model regresi semiparametrik campuran spline *truncated* dan kernel ?
2. Bagaimana mengaplikasikan estimasi interval terpendek untuk parameter model regresi semiparametrik campuran spline *truncated* dan kernel pada data Tingkat Pengangguran Terbuka ?

## **1.3 Tujuan Penelitian**

Berdasarkan permasalahan di atas, tujuan dari penelitian ini adalah sebagai berikut.

1. Mengkaji estimasi interval terpendek untuk parameter komponen parametrik dan spline dalam regresi semiparametrik campuran spline *truncated* dan kernel.
2. Mengaplikasikan estimasi interval terpendek untuk parameter model regresi nonparametrik campuran spline *truncated* dan kernel pada data Tingkat Pengangguran Terbuka (TPT) di Indonesia tahun 2016.

## **1.4 Manfaat Penelitian**

Manfaat yang ingin dicapai dari hasil penelitian ini adalah sebagai berikut:

1. Menambah wawasan keilmuan yang lebih khusus kepada penulis tentang estimasi interval parameter dalam regresi semiparametrik campuran spline dan kernel.
2. Memberikan informasi yang lebih rinci mengenai Tingkat Pengangguran Terbuka (TPT) sehingga dapat membantu pemerintah dalam penentuan kebijakan di bidang sosial ekonomi, terutama penanggulangan masalah pengangguran.
3. Memberikan alternatif bagi Badan Pusat Statistik (BPS) mengenai metode analisis data khususnya untuk data yang sesuai dengan regresi semiparametrik campuran spline dan kernel.

### **1.5 Batasan Masalah**

Berdasarkan rumusan masalah, maka yang menjadi batasan permasalahan dalam penelitian ini antara lain:

1. Fungsi yang digunakan dalam estimator campuran adalah fungsi Spline *Truncated* linier dan kernel Nadaraya-Watson dengan fungsi *Gaussian*.
2. Metode yang digunakan untuk menentukan titik knot dan *bandwidth* optimum menggunakan metode *Generalized Cross Validation (GCV)*.
3. Data yang digunakan dalam penelitian ini adalah data Tingkat Pengangguran Terbuka menurut Provinsi Tahun 2016 yang bersumber dari publikasi BPS.

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Regresi Parametrik

Regresi parametrik digunakan jika penyebaran pola data membentuk pola tertentu atau diketahui. Dengan mengetahui bentuk kurva atau fungsi regresi, maka analisis regresi parametrik akan lebih menguntungkan, khususnya metode inferensi dan interpretasi parameternya akan lebih sederhana. Oleh karena itu analisis regresi parametrik lebih sering digunakan apabila terdapat informasi tentang bentuk kurva regresinya (Eubank, 1999). Dilihat dari pandangan statistik inferensia, pendekatan model regresi parametrik memiliki sifat yang sangat baik, seperti misalnya sederhana, mudah interpretasinya, parsimoni, tak bias, estimator linier, efisien, konsisten, dan *Best Linier Unbiased Estimator* (BLUE) yang tidak dimiliki oleh pendekatan model regresi nonparametrik (Budiantara, 2009).

Secara matematis fungsi regresi parametrik bisa ditulis dengan persamaan sebagai berikut:

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

fungsi  $f(x_i)$  seringkali disebut sebagai fungsi regresi parametrik atau kurva regresi parametrik yang memiliki *error* random  $\varepsilon_i$  dan diasumsikan berdistribusi normal independen dengan rata-rata nol dan variansi  $\sigma^2$  (Eubank, 1999). Fungsi  $f(x_i)$  dapat dituliskan dalam bentuk :

$$y_i = \tilde{x}_i' \tilde{\beta} + \varepsilon_i$$

dimana  $\tilde{x}_i' = [1 \quad x_{i1} \quad x_{i2} \dots x_{ip}]$ ,  $i = 1, 2, \dots, n$ ; sedangkan  $n$  adalah banyak data

dan  $p$  adalah banyak variabel, sementara  $\tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ . Sehingga persamaannya

menjadi:

$$= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.1)$$

Model regresi (2.1) dapat ditulis dalam bentuk matrik :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

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

dimana  $\tilde{\mathbf{y}}$  merupakan vektor dari variabel respon berukuran  $n \times 1$ ,  $\mathbf{X}$  merupakan matrik berukuran  $n \times (p + 1)$  dan  $\tilde{\boldsymbol{\beta}}$  adalah vektor parameter yang akan diestimasi berukuran  $(p + 1) \times 1$ ,  $\tilde{\boldsymbol{\varepsilon}}$  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  $\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}$  sehingga diperoleh :

$$\hat{\tilde{\boldsymbol{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{y}} \quad (2.3)$$

## 2.2 Regresi Nonparametrik 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  $y_i$  adalah variabel respon 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:

$$y_i = g(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.4)$$

Selanjutnya jika kurva regresi  $g(t_i)$  pada persamaan (2.4) dihampiri dengan kurva regresi spline *truncated* dengan knot  $K_1, K_2, \dots, K_r$  maka:

$$g(t_i) = \sum_{j=0}^m \theta_j t_i^j + \sum_{k=1}^r \theta_{k+m} (t_i - K_k)_+^m \quad (2.5)$$

dengan  $\theta$  adalah parameter-parameter model dan fungsi *truncated*  $(t_i - K_k)$  diberikan oleh :

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

sehingga model regresi pada persamaan (2.4) menjadi :

$$y_i = \sum_{j=0}^m \theta_j t_i^j + \sum_{k=1}^r \theta_{k+m} (t_i - K_k)_+^m + \varepsilon_i \quad (2.6)$$

atau dapat ditulis dalam notasi matrik :

$$\tilde{y} = G(\mathbf{k})\tilde{\theta} + \tilde{\varepsilon} \quad (2.7)$$

dimana :

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{G}(\mathbf{k}) = \begin{bmatrix} 1 & t_1^1 & \dots & t_1^m & (t_1 - K_1)_+^m & (t_1 - K_2)_+^m & \dots & (t_1 - K_r)_+^m \\ 1 & t_2^1 & \dots & t_2^m & (t_2 - K_1)_+^m & (t_2 - K_2)_+^m & \dots & (t_2 - K_r)_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n^1 & \dots & t_n^m & (t_n - K_1)_+^m & (t_n - K_2)_+^m & \dots & (t_n - K_r)_+^m \end{bmatrix}$$

$$\tilde{\theta} = [\theta_0 \quad \theta_1 \quad \dots \quad \theta_m \quad \theta_{m+1} \quad \theta_{m+2} \quad \dots \quad \theta_{m+r}]^T, \text{ dan } \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

Kurva regresi  $g(t_i)$  merupakan kurva regresi nonparametrik spline *truncated* derajat  $m$  dengan banyaknya titik knot  $r$ . Derajat  $m$  merupakan derajat pada persamaan polinomial. Titik-titik knot  $K_1, K_2, \dots, K_r$  adalah titik-titik knot yang menunjukkan perubahan pola perilaku dari kurva pada sub-sub interval yang berbeda, dimana  $K_1 < K_2 < \dots < K_r$ .

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

$$\hat{\theta} = [G(\mathbf{k})'G(\mathbf{k})]^{-1}G(\mathbf{k})'\tilde{y} \quad (2.8)$$

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

$$\begin{aligned}\hat{g}(t_i) &= G(\mathbf{k})\hat{\boldsymbol{\theta}} \\ &= G(\mathbf{k})[G(\mathbf{k})'G(\mathbf{k})]^{-1}G(\mathbf{k})'\tilde{\mathbf{y}} \\ &= A(\mathbf{k})\tilde{\mathbf{y}}\end{aligned}\quad (2.9)$$

dimana

$$A(k) = G(k)[G(k)'G(k)]^{-1}G(k)'$$

### 2.3 Regresi Nonparametrik Kernel

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

$$\mathbf{y}_i = \mathbf{h}(z_i) + \boldsymbol{\varepsilon}_i \quad (2.10)$$

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

Menurut Hardle (1994), kurva  $h(z_i)$  pada model nonparametrik dapat diestimasi dengan pendekatan kernel yang didasarkan pada fungsi densitas kernel. Kurva  $h(z_i)$  pada model (2.10) dapat diestimasi menggunakan estimator kernel Nadaraya-Watson.

$$\hat{h}_\alpha(z) = \mathbf{n}^{-1} \sum_{i=1}^n \left[ \frac{K_\alpha(z - z_i)}{\mathbf{n}^{-1} \sum_{j=1}^n K_\alpha(z - z_j)} \right] \mathbf{y}_i = \mathbf{n}^{-1} \sum_{i=1}^n W_{\alpha i}(z) \mathbf{y}_i \quad (2.11)$$

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

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

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

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

Estimator kernel sangat bergantung pada fungsi kernel K dan pemilihan *bandwidth*  $\alpha$ . Dalam bentuk matrik dapat dituliskan :

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

sehingga persamaan (2.11) dapat dituliskan dalam notasi matrik seperti dibawah ini.

$$\hat{\mathbf{h}}_\alpha(\mathbf{z}) = \mathbf{D}(\alpha) \tilde{\mathbf{y}} \quad (2.12)$$

dimana vector  $\hat{\mathbf{h}}_\alpha(\mathbf{z})$  berukuran  $n \times 1$ , vektor  $\tilde{\mathbf{y}}$  berukuran  $n \times 1$  dan matrik  $\mathbf{D}(\alpha)$  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  $\alpha$  adalah *bandwidth*. Estimator kernel sangat bergantung pada pemilihan *bandwidth*  $\alpha$  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).



## 2.4 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.

Misalkan diberikan data berpasangan  $(x_i, t_i, z_i, y_i)$ ,  $i = 1, 2, \dots, n$  maka model semiparametrik dirumuskan sebagai berikut :

$$y_i = \mu(x_i, t_i, z_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.13)$$

dengan kurva diasumsikan aditif, maka diperoleh :

$$y_i = f(x_i) + g(t_i) + h(z_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.14)$$

dimana  $y_i$  adalah variabel respon,  $f(x_i)$  adalah komponen parametrik,  $g(t_i)$  adalah komponen nonparametrik spline,  $h(z_i)$  adalah komponen nonparametrik kernel, dan  $\varepsilon_i$  adalah error random yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian  $\sigma^2$ .

Kurva regresi komponen parametrik  $f(x_i)$  pada persamaan (2.14) dihampiri dengan fungsi linier (2.1) dan dapat dituliskan dalam bentuk matrik (2.2). Selanjutnya kurva regresi  $g(t_i)$  dihampiri dihampiri dengan fungsi spline (2.6) dan dituliskan dalam bentuk matrik (2.7). Kurva regresi  $h(z_i)$  dihampiri dengan fungsi kernel (2.11) dan dapat dituliskan dalam bentuk matrik (2.12).

Model regresi semiparametrik pada persamaan (2.14) dapat disajikan dalam bentuk matrik sebagai berikut.

$$\tilde{y} = X\tilde{\beta} + G(\mathbf{k})\tilde{\theta} + D(\boldsymbol{\alpha})\tilde{\gamma} + \tilde{\varepsilon} \quad (2.15)$$

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  $(p + 1) \times 1$ ,  $\mathbf{G}$  merupakan matrik komponen spline berukuran  $n \times (r + m)$ , vektor  $\tilde{\theta}$

berukuran  $(r + m) \times 1$ , matrik  $D(\alpha)$  berukuran  $n \times n$  dan  $\tilde{\varepsilon}$  adalah vektor *error* random berukuran  $n \times 1$

## 2.5 Generalized Cross Validation (GCV)

Hal penting yang berperan dalam mendapatkan estimator spline *truncated* terbaik adalah pemilihan titik knot yang optimal. Salah satu metode yang sering digunakan dalam memilih titik knot optimal adalah *Generalized Cross Validation* (GCV). Jika dibandingkan dengan metode lain, misalnya *Cross Validation* (CV) dan metode *Unbiased Risk* (UBR) ataupun *Generalized Maximum Likelihood* (GML), GCV secara teoritis memiliki sifat optimal asimtotik. Metode GCV juga memiliki kelebihan tidak memerlukan pengetahuan terhadap variansi populasi  $\sigma^2$  serta metode GCV invariants terhadap transformasi. Metode GCV merupakan pengembangan dari CV (Wahba, 1990).

Sementara itu pada kernel bergantung pada pemilihan *bandwidth*. *Bandwidth*  $\alpha$  adalah parameter pemulus (*smoothing*) yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. Titik knot maupun bandwidth yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan fluktuatif, sebaliknya titik knot maupun *bandwidth* yang terlalu besar atau lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data (Hardle, 1994). Pemilihan titik knot  $k$  dan bandwidth yang optimum menggunakan GCV di definisikan sebagai berikut.

$$GCV(\lambda, \alpha) = \frac{MSE(k, \alpha)}{(n^{-1} \text{trace}[(I - M(k, \alpha))]^2)} \quad (2.16)$$

dimana :

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

## 2.6 Estimasi Interval

Estimasi Interval merupakan pengembangan dari estimasi titik, dimana nilai estimasi tidak bergantung pada satu titik akan tetapi memuat suatu range tertentu yang memiliki batas bawah dan batas atas. Dibandingkan dengan estimasi

titik, estimasi interval dapat memberikan informasi dengan akurasi yang lebih baik dalam menentukan estimasi suatu parameter populasi. Oleh karena itu dirumuskanlah suatu interval random, yang merupakan interval suatu parameter populasi misalnya  $\beta$ , yang batas-batasnya merupakan statistik. Interval ini disusun sedemikian hingga memiliki peluang yang sebesar mungkin. Montgomery et al (2012) menjelaskan tahapan untuk mendapatkan estimasi interval untuk suatu parameter dalam regresi dengan menggunakan metode *pivotal quantity* sebagai berikut.

Model regresi linier berganda seperti yang disajikan pada persamaan (2.1), diketahui bahwa  $y_i$  berdistribusi normal dengan mean  $\beta_0 + \sum_{j=1}^l \beta_j x_j$  dan varians  $\sigma^2$ . Langkah pertama yang harus dilakukan yaitu mengetahui estimasi parameter, dalam hal ini pada metode regresi linier berganda digunakan metode *Ordinary Least Square* (OLS). Mengacu pada persamaan (2.2) diperoleh estimator  $\hat{\beta}$  dengan metode OLS (Kutner, et.al., 2004) seperti dibawah ini.

$$\hat{\beta} = (X^T X)^{-1} X^T \tilde{y} \quad (2.17)$$

Dengan memperhatikan bahwa  $\hat{\beta}$  kombinasi linier dari  $\tilde{y}$ , maka  $\hat{\beta} \sim N(\tilde{\beta}, \sigma^2 (X^T X)^{-1})$ , sehingga untuk masing-masing estimator berdistribusi normal  $\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{jj})$ , dimana  $C_{jj}$  merupakan elemen diagonal ke- $j$  dari matriks  $(X^T X)^{-1}$ . Akibatnya, diperoleh *pivotal quantity*

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}}, j = 0, 1, \dots, p \quad (2.18)$$

berdistribusi  $t$  dengan derajat bebas  $n-p$ , dimana  $\sigma^2$  diestimasi dengan varians *error*. Berdasarkan hasil pada persamaan (2.18), didapat interval konfidensi terpendek berukuran  $(1 - \alpha) \times 100\%$  untuk koefisien regresi  $\beta_j, j = 0, 1, \dots, p$  adalah sebagai berikut.

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\sigma^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\sigma^2 C_{jj}} \quad (2.19)$$

## 2.7 Pivotal Quantity

*Pivotal quantity* atau besaran pivot merupakan kuantitas yang distribusinya tidak tergantung atau tidak memuat parameter. Misalkan diberikan suatu variabel random  $X = (x_1, x_2, \dots, x_n)$ , dan  $\theta$  adalah parameter populasi, maka suatu statistik  $T(x_1, x_2, \dots, x_n, \theta)$  dikatakan *Pivotal quantity* untuk parameter  $\theta$ , jika distribusi dari  $T(x_1, x_2, \dots, x_n, \theta)$  tidak memuat atau tidak bergantung pada parameter  $\theta$ .

*Pivotal quantity* sangat penting dalam konstruksi statistik uji, dengan menggunakan prinsip dasar bahwa suatu statistik tidak memuat parameter dari populasinya, seperti konstruksi statistik t-student untuk kasus distribusi normal dengan varians populasi tidak diketahui.

*Pivotal quantity* juga dapat digunakan untuk mengkonstruksi estimasi interval. Dimisalkan ingin dicari estimasi interval untuk parameter  $\theta$ , maka langkah awal untuk mendapatkan estimasi interval yaitu dengan mencari estimasi titik dari parameter  $\theta$ , kemudian didapatkan distribusi sampling dari estimasi titik tersebut. Langkah selanjutnya dicari *pivotal quantity* untuk parameter  $\theta$  dan dilanjutkan dengan mencari estimasi interval terpendek dengan menyelesaikan persamaan probabilitas seperti dibawah ini.

$$P(a(x_1, x_2, \dots, x_n) \leq T(x_1, x_2, \dots, x_n, \theta) \leq b(x_1, x_2, \dots, x_n)) = 1 - \alpha \quad (2.20)$$

dimana  $T(x_1, x_2, \dots, x_n, \theta)$  merupakan pivotal quantity untuk parameter  $\theta$ . Selanjutnya akan didapatkan estimasi interval  $1 - \alpha$  untuk parameter  $\theta$  sedemikian hingga dapat dituliskan dalam bentuk persamaan :

$$P(u_1(x_1, x_2, \dots, x_n) \leq \theta \leq u_2(x_1, x_2, \dots, x_n)) = 1 - \alpha \quad (2.21)$$

dimana  $u_1(x_1, x_2, \dots, x_n)$  merupakan batas atas, dan  $u_2(x_1, x_2, \dots, x_n)$  merupakan batas bawah (Widiarini, 2009).

## 2.8 Koefisien Determinasi ( $R^2$ )

Salah satu tujuan analisis regresi adalah memperoleh model terbaik yang mampu menjelaskan hubungan antara variabel prediktor dan variabel respon. Salah

satu kriteria yang dapat digunakan untuk memilih model terbaik adalah dengan menggunakan koefisien determinasi  $R^2$ . Koefisien determinasi ( $R^2$ ) merupakan ukuran ketelitian atau ketepatan model regresi, dapat juga diartikan besarnya kontribusi prediktor terhadap respon. Rumus koefisien determinasi sebagai berikut.

$$R^2 = \frac{SSR}{SST} = \frac{(\hat{\mathbf{y}} - \bar{\mathbf{y}})'(\hat{\mathbf{y}} - \bar{\mathbf{y}})}{(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})} \quad (2.22)$$

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

## 2.9 Teorema Dasar Terkait Aljabar Matriks

Untuk mendapatkan estimasi interval untuk parameter model regesi semiparametrik Spline linier berikut ini diberikan beberapa teorema aljabar matriks. (Rencher dan Schaalje,2008).

### 1. Teorema 2.11 (Rencher dan Schaalje,2008)

Apabila  $\mathbf{A}$  suatu matriks berukuran  $n \times p$  dan  $\mathbf{B}$  berukuran  $p \times n$ , maka

$$tr(\mathbf{AB}) = tr(\mathbf{BA}).$$

### 2. Teorema 2.13d (Rencher dan Schaalje,2008)

Jika matriks  $\mathbf{A}$  merupakan matriks simetris dan idempotent dengan rank  $r$ , maka  $rank(\mathbf{A}) = trace(\mathbf{A})$ .

### 3. Teorema 2.14a (Rencher dan Schaalje,2008)

Diberikan vektor  $\tilde{\mathbf{a}}$  dan  $\tilde{\mathbf{x}}$ , dimana  $\tilde{\mathbf{a}}'\tilde{\mathbf{x}} = \tilde{\mathbf{x}}'\tilde{\mathbf{a}}$ ,  $\tilde{\mathbf{a}}' = (a_1, a_2, \dots, a_p)$  memuat konstanta, dengan ini berlaku

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

### 4. Teorema 2.14b (Rencher dan Schaalje,2008)

Apabila vektor  $\tilde{\mathbf{x}}$  dan  $\mathbf{A}$  merupakan suatu matriks simetris, maka

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

**5. Teorema 4.4a** (Rencher dan Schaalje,2008)

Misalkan  $\tilde{y}$  vector random berukuran  $p \times 1$  dan berdistribusi  $N_p(\tilde{\mu}, \Sigma)$  dan  $\tilde{a}$  merupakan vector konstanta berukuran  $p \times 1$  serta  $\mathbf{A}$  merupakan matriks konstanta nonsingular berukuran  $k \times p$  dengan  $\text{rank } k \leq p$ , maka distribusi kombinasi linier dari  $\tilde{a}'\tilde{y}$  adalah  $\tilde{a}'\tilde{y} \sim N(\tilde{a}'\tilde{\mu}, \tilde{a}'\Sigma\tilde{a})$  serta distribusi dari kombinasi linier  $\mathbf{A}\tilde{y}$  adalah  $\mathbf{A}\tilde{y} \sim N(\mathbf{A}\tilde{\mu}, \mathbf{A}\Sigma\mathbf{A}')$ .

**6. Teorema 5.5** (Rencher dan Schaalje,2008)

Misalkan  $\tilde{y}$  vektor random berdistribusi  $N_p(\tilde{\mu}, \Sigma)$ ,  $\mathbf{A}$  matriks simetris dengan  $\text{rank}(\mathbf{A}) = r$  dan untuk  $\lambda = \frac{1}{2} \tilde{\mu}'\mathbf{A}\tilde{\mu}$  maka  $\tilde{y}'\mathbf{A}\tilde{y}$  akan mengikuti distribusi  $\chi^2_{(r, \lambda)}$  jika dan hanya jika  $\mathbf{A}\Sigma$  idempoten.

**Akibat 1.**

Jika  $\tilde{y} \sim N_p(\tilde{0}, \mathbf{I})$  maka  $\tilde{y}'\mathbf{A}\tilde{y}$  akan mengikuti distribusi  $\chi^2_{(r)}$  jika dan hanya jika  $\mathbf{A}$  idempotent dengan  $\text{rank}(\mathbf{A}) = r$ .

**Akibat 2.**

Jika  $\tilde{y} \sim N(\tilde{\mu}, \sigma^2 I)$  dan  $\mathbf{A}$  matriks simetris dengan  $\text{rank}(\mathbf{A}) = r$ , maka  $\tilde{y}'\mathbf{A}\tilde{y} / \sigma^2$  berdistribusi  $\chi^2(r, \tilde{\mu}'\mathbf{A}\tilde{\mu} / 2\sigma^2)$  jika dan hanya jika  $\mathbf{A}$  idempotent.

**7. Teorema 5.6b** (Rencher dan Schaalje,2008)

Misalkan  $\mathbf{B}$  matriks berukuran  $k \times p$  dan  $\mathbf{A}$  merupakan matriks simetris berukuran  $p \times p$  serta  $\tilde{y}$  berdistribusi  $N_p(\tilde{\mu}, \Sigma)$  maka  $\mathbf{B}\tilde{y}$  dan  $\tilde{y}'\mathbf{A}\tilde{y}$  dikatakan saling independen jika dan hanya jika  $\mathbf{B}\Sigma\mathbf{A} = 0$ .

**Akibat 1.**

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

**8. Definisi 2.14.3 Metode Pengali Lagrange** (Rencher dan Schaalje,2008)

Jika diberikan  $u = f(\tilde{x})$  dengan  $\tilde{x} = (x_1, x_2, \dots, x_p)'$  dan diberikan suatu kendala  $h_1(x) = 0, h_2(x) = 0, h_3(x) = 0, \dots, h_q(x) = 0$  atau dapat ditulis  $\tilde{h}(\tilde{x}) = 0$ , sehingga dapat dituliskan  $v = u + \tilde{\lambda}'\tilde{h}(\tilde{x}) = 0$  dengan

$\tilde{\lambda}' = (\lambda_1, \lambda_2, \dots, \lambda_q)$  maka untuk  $\tilde{y}' = (\tilde{x}', \tilde{\lambda}')$ , nilai maksimum atau minimum dari  $u = f(\tilde{x})$  dengan kendala  $\tilde{h}(\tilde{x}) = 0$  dipenuhi oleh syarat

$$\frac{\partial v}{\partial \tilde{y}} = \tilde{0}$$

Atau ekuivalen dengan

$$\frac{\partial v}{\partial \tilde{x}} + \frac{\partial \tilde{h}}{\partial \tilde{x}} \tilde{\lambda} = \tilde{0} \text{ dan } \tilde{h}(\tilde{x}) = 0.$$

### 9. Definisi 5.42a Distribusi-t NonSentral (Rencher dan Schaalje,2008)

i) Jika  $M_1 \sim N(0,1)$  ,  $M_2 \sim \chi^2_{(p)}$  dan  $M_1$  dan  $M_2$  independen maka

$$T = \frac{M_1}{\sqrt{\frac{M_2}{p}}} \sim t_{(p)}.$$

ii) Jika  $M_1 \sim N(\mu,1)$  ,  $M_2 \sim \chi^2_{(p)}$  dan  $M_1$  dan  $M_2$  independen maka

$$T = \frac{M_1}{\sqrt{\frac{M_2}{p}}} \sim t_{(p,\mu)}.$$

iii) Jika  $M_1 \sim N(\mu,\sigma^2)$  ,  $M_2 \sim \chi^2_{(p)}$  dan  $M_1$  dan  $M_2$  independen maka

$$T = \frac{\frac{M_1}{\sigma}}{\sqrt{\frac{M_2}{p}}} \sim t_{\left(p, \frac{\mu}{\sigma}\right)}.$$

### 2.10Tingkat Pengangguran Terbuka (TPT)

Tingkat Pengangguran merupakan salah satu indikator yang digunakan sebagai salah satu alat ukur dalam bidang ketenagakerjaan. Konsep definisi ketenagakerjaan pada penelitian ini menggunakan konsep Badan Pusat Statistik (BPS) yang merujuk pada rekomendasi ILO, seperti tercantum dalam buku “Surveys of Economically Active Population, Employment, Unemployment and Under employment : An ILO Manual on Concepts and Methods”, ILO 1992. Sehingga data ketenagakerjaan yang dihasilkan dari berbagai survei di Indonesia

dapat dibandingkan secara internasional, tanpa mengesampingkan kondisi ketenagakerjaan spesifik Indonesia.

Berdasarkan definisi dari BPS, angkatan kerja diartikan sebagai penduduk usia kerja (15 tahun ke atas) yang bekerja, atau punya pekerjaan namun sementara tidak bekerja dan pengangguran. Sementara itu yang bukan termasuk angkatan kerja adalah penduduk usia kerja (15 tahun ke atas) yang masih sekolah, mengurus rumah tangga, atau melaksanakan kegiatan lainnya.

Tingkat Pengangguran Terbuka (TPT) sendiri didefinisikan sebagai persentase jumlah pengangguran terhadap jumlah angkatan kerja. Pengangguran terbuka terdiri dari:

- mereka yang tak punya pekerjaan dan mencari pekerjaan,
- mereka yang tak punya pekerjaan dan mempersiapkan usaha,
- mereka yang tak punya pekerjaan dan tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan, dan
- mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja.



Gambar 2.1 Bagan Konsep Ketenagakerjaan

TPT mengindikasikan besarnya persentase angkatan kerja yang termasuk dalam pengangguran. Variabel untuk menyusun indikator ini diperoleh dari Survei Angkatan Kerja Nasional (Sakernas). TPT yang tinggi menunjukkan bahwa



terdapat banyak angkatan kerja yang tidak terserap pada pasar kerja. Misalkan terdapat TPT 5%, artinya dari 100 penduduk usia 15 tahun keatas yang tersedia untuk memproduksi barang dan jasa (angkatan kerja) sebanyak 5 orang merupakan pengangguran.

## **2.11 Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka**

Penelitian terkait TPT telah banyak dilakukan, baik di negara berkembang seperti Indonesia maupun di negara maju. Berikut ini adalah beberapa faktor yang diduga berpengaruh terhadap TPT baik menurut ahli maupun hasil dari penelitian yang telah dilakukan.

Dilihat dari aspek pendidikan, menurut Elhorst (2003) dalam beberapa studi menyatakan pendidikan berpengaruh secara signifikan terhadap TPT. Seseorang yang mempunyai pendidikan tinggi memiliki ketrampilan yang dibutuhkan oleh perekonomian dengan perkembangan teknologi yang terus meningkat. Seseorang dengan pendidikan yang rendah sangat sulit mendapatkan pekerjaan yang layak. Dampak dengan rendahnya pendidikan maka ketrampilan yang dimiliki terbatas. Selain itu terdapat ketidakcocokan ketrampilan yang dibutuhkan antara pencari tenaga kerja dengan lapangan kerja yang dibuka dan budaya masyarakat yang belum berorientasi kepada daya saing dan produktivitas. Hal ini sejalan dengan Human Capital (teori modal manusia) yang menyatakan bahwa pendidikan termasuk kedalam salah satu investasi pada bidang sumber daya manusia (Idris, 2007). Investasi pendidikan merupakan kegiatan yang dapat dinilai sebagai stok manusia, dimana nilai stok manusia setelah mengikuti berbagai jenis dan bentuk pendidikan diharapkan dapat memperoleh peningkatan penghasilan individu, peningkatan produktivitas kerja, dan peningkatan nilai rasional (*social benefit*). Salah satu variabel indikator pendidikan yang dapat digunakan adalah persentase penduduk 15 tahun keatas yang lulus SMA (Astuti dkk, 2017).

Jika dilihat dari aspek ekonomi, pertumbuhan ekonomi akan mempunyai dampak langsung terhadap penciptaan kesempatan kerja, sehingga akan mengatasi pengangguran (Hasan, 2005). Pertumbuhan ekonomi bisa memberikan peluang kesempatan kerja baru ataupun memberikan kesempatan industri untuk meningkatkan output yang berdampak pada peningkatan penggunaan faktor

produksi dan akan meningkatkan kemakmuran masyarakat (Sukirno, 2000), salah satunya yaitu tenaga kerja, sehingga dapat mengurangi jumlah pengangguran. Alghofari (2010) dan Tantri dan Ratnasari (2016) juga menyatakan dalam penelitiannya bahwa pertumbuhan ekonomi mempunyai kecenderungan hubungan negatif dan kuat terhadap tingkat pengangguran. Hal ini sesuai Hukum Okun yang menyatakan hubungan negatif antara pertumbuhan ekonomi dengan pengangguran (Mankiw, 2006).

Investasi sebagai salah satu pendorong perekonomian juga ikut berpengaruh terhadap pengangguran secara tidak langsung. Terciptanya lapangan usaha baru dan menguatnya kegiatan ekonomi sebagai akibat dari iklim investasi yang baik dapat mengurangi pengangguran. Dalam penelitiannya (Sari, 2012) menyatakan bahwa investasi berpengaruh terhadap jumlah pengangguran. Hubungan antara Investasi dengan pengangguran dapat dilihat berdasarkan Teori Harrod Domar (Mulyadi, 2003), dalam teorinya berpendapat bahwa investasi tidak hanya menciptakan permintaan, tetapi juga memperbesar kapasitas produksi. Artinya dengan semakin besar kapasitas produksi maka akan membutuhkan tenaga kerja yang semakin besar pula dengan asumsi full employment. Ini karena investasi merupakan penambahan faktor-faktor produksi, yang mana salah satu dari faktor produksi adalah tenaga kerja. Dengan begitu perekonomian secara keseluruhan dapat menyerap tenaga kerja yang sebanyak-banyaknya, sehingga partisipasi angkatan kerja akan semakin meningkat pula.

Besaran upah minimum juga dapat memberikan pengaruh terhadap pengangguran yang ada di Indonesia. Kenaikan upah minimum yang signifikan secara langsung berdampak terhadap biaya tenaga kerja yang dikeluarkan oleh badan usaha, akibatnya akan ada efisiensi tenaga kerja yang dilakukan. Menurut (Mankiw, 2003) adanya kekakuan upah (*wage rigidity*) yakni gagalnya upah melakukan penyesuaian sampai penawaran tenaga kerja sama dengan permintaannya. Upah tidak selalu fleksibel menyeimbangkan penawaran dan permintaan tenaga kerja, kadang-kadang upah riil tertahan di atas tingkat kliring-pasar. Ketika upah riil di atas tingkat yang menyeimbangkan penawaran dan permintaan, jumlah tenaga kerja yang ditawarkan melebihi jumlah yang diminta sehingga terjadi pengangguran. Hal ini sejalan dengan penelitian yang dilakukan

Sari (2012), bahwa investasi dan besaran upah minimum berpengaruh secara signifikan terhadap jumlah pengangguran. Setiap kenaikan tingkat upah akan diikuti oleh turunnya tenaga kerja yang diminta, yang berarti akan menyebabkan bertambahnya pengangguran (Alghofari, 2010).

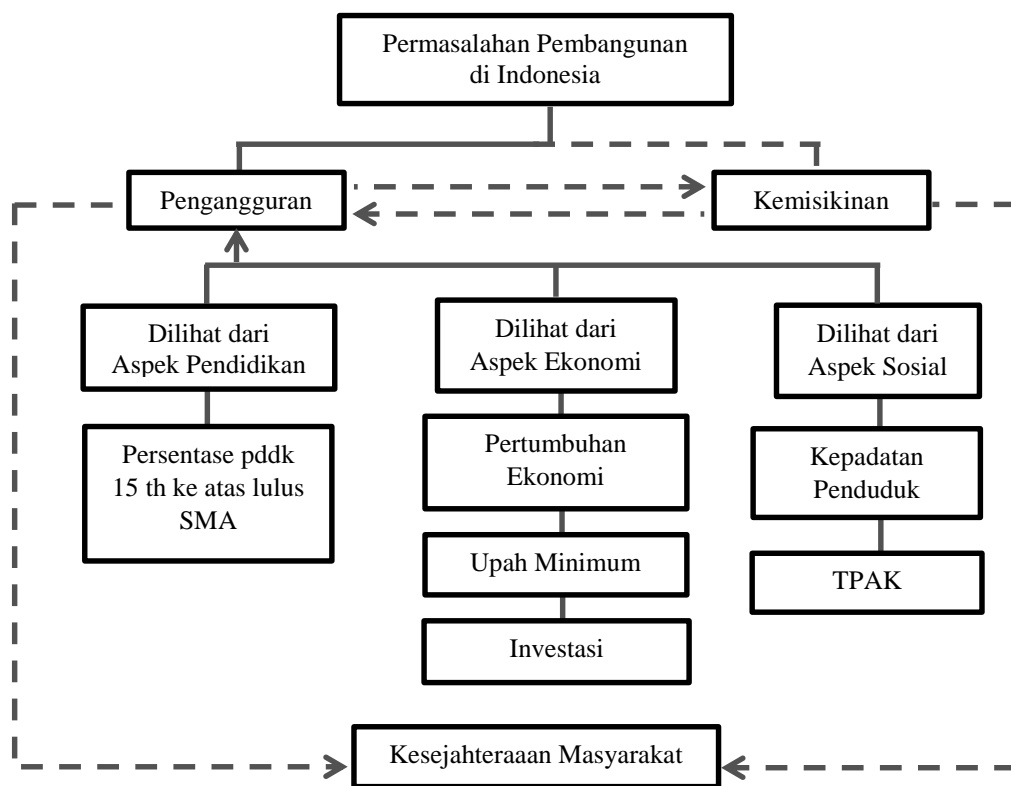
Selain itu jika dilihat dari aspek sosial, kepadatan penduduk dan partisipasi angkatan kerja ikut memberikan dampak terhadap TPT. Kepadatan penduduk merupakan perbandingan antara jumlah penduduk dengan luas wilayahnya. Semakin tinggi kepadatan penduduk bisa mengakibatkan munculnya permasalahan baru di wilayah tersebut, seperti kemiskinan, pengangguran, lingkungan kumuh dan persoalan sosial ekonomi lainnya. Status perkotaan atau pedesaan cenderung berpengaruh terhadap tingkat pengangguran. Pada umumnya tingkat pengangguran di perkotaan lebih tinggi jika dibandingkan dengan pedesaan (Setiawan dalam Faidah, 2012). Frenkle dan Ros (2006) menyebutkan bahwa pertumbuhan penduduk berhubungan dengan tingkat pengangguran. Dalam penelitiannya, Ramadhani (2016) juga menyatakan bahwa kepadatan penduduk berpengaruh secara signifikan terhadap tingkat pengangguran.

Tingkat Partisipasi Angkatan Kerja (TPAK) merupakan suatu ukuran proporsi penduduk usia kerja yang terlibat secara aktif dalam pasar tenaga kerja baik yang bekerja maupun sedang mencari pekerjaan. Indikator ini dapat mengindikasikan besaran ukuran relatif penawaran tenaga kerja (supply) yang dapat terlibat dalam produksi barang dan jasa. Pertumbuhan penduduk yang tinggi akan menyebabkan semakin meningkatnya jumlah tenaga kerja. Jika tidak diimbangi dengan penambahan kesempatan kerja akan menyebabkan bertambahnya jumlah pengangguran. Penduduk yang tidak mendapatkan kesempatan kerja tersebut tidak bisa ikut berpartisipasi dalam kegiatan perekonomian sehingga terjadilah pengangguran. Astuti (2017) menyatakan dalam penelitiannya bahwa TPAK berpengaruh terhadap TPT.

Sejalan dengan teori Malthus yang apabila ditelaah lebih dalam menyatakan penduduk cenderung bertumbuh secara tak terbatas hingga mencapai batas persediaan makanan, dalam hal ini menimbulkan manusia saling bersaing dalam menjamin kelangsungan hidupnya dengan cara mencari sumber makanan, dengan persaingan ini maka akan ada sebagian manusia yang tersisih serta tidak

mampu lagi memperoleh bahan makanan. Pada masyarakat modern diartikan bahwa semakin pesatnya jumlah penduduk akan menghasilkan tenaga kerja yang semakin banyak pula, namun hal ini tidak diimbangi dengan kesempatan kerja yang ada. Karena jumlah kesempatan yang sedikit itulah maka manusia saling bersaing dalam memperoleh pekerjaan dan yang tersisih dalam persaingan tersebut menjadi golongan penganggur.

Berdasarkan beberapa faktor di atas dibuatlah kerangka konseptual seperti Gambar 2.2 sebagai berikut.



Gambar 2.2 Diagram Konseptual Penelitian

Dengan berdasarkan kerangka konseptual yang digunakan serta beberapa penelitian yang melakukan kajian tentang determinan pengangguran, maka pada penelitian ini akan melibatkan aspek pendidikan, aspek sosial dan aspek ekonomi. Aspek pendidikan yang digunakan adalah Persentase penduduk 15 tahun ke atas lulus SMA/SMK. Sedangkan aspek sosial yang digunakan adalah rata-rata kepadatan jumlah penduduk dan tingkat partisipasi angkatan kerja (TPAK). Untuk

merepresentasikan aspek ekonomi variabel yang digunakan adalah laju pertumbuhan ekonomi, upah minimum provinsi dan investasi dalam negeri.

## BAB 3

### METODOLOGI PENELITIAN

#### 3.1 Sumber Data

Penelitian ini menggunakan data sekunder yang berasal dari publikasi yang diterbitkan oleh BPS yaitu Statistik Indonesia 2017 dan Statistik Kesejahteraan Rakyat 2016. Unit observasi yang digunakan dalam penelitian ini adalah seluruh provinsi yang ada di Indonesia, yaitu sebanyak 34 provinsi.

#### 3.2 Variabel Penelitian

Dalam penelitian ini variabel respon dan variabel-variabel prediktor yang digunakan adalah sebagaimana Tabel 3.1.

Tabel 3.1 Variabel Penelitian

|       | Variabel  | Skala Data | Satuan               |
|-------|---|------------|----------------------|
| $y$   | : TPT (Tingkat Pengangguran Terbuka tingkat provinsi di Indonesia Tahun 2016) | Rasio      | Persen               |
| $x_1$ | : Tingkat Partisipasi Angkatan Kerja  | Rasio      | Persen               |
| $x_2$ | : Upah Minimum Provinsi   | Rasio      | Rupiah               |
| $x_3$ | : Persentase penduduk usia 15+ yang lulus SMA/SMK                             | Rasio      | Persen               |
| $x_4$ | : Kepadatan Penduduk  | Rasio      | jiwa/km <sup>2</sup> |
| $x_5$ | : Investasi   | Rasio      | Triliun<br>Rupiah    |
| $x_6$ | : Laju Pertumbuhan Ekonomi  | Rasio      | Persen               |

#### 3.3 Struktur Data

Dalam penelitian ini struktur data yang digunakan adalah data dengan variabel respon dan variabel prediktor terdiri atas komponen parametrik dan nonparametrik.

Tabel 3.2 Struktur Data Penelitian

| Observasi | Respon |             | Prediktor   |             |             |             |             |
|-----------|--------|-------------|-------------|-------------|-------------|-------------|-------------|
|           | $y$    | $x_1$       | $x_2$       | $x_3$       | $x_4$       | $x_5$       | $x_6$       |
| 1         | $y_1$  | $x_{1(1)}$  | $x_{2(1)}$  | $x_{3(1)}$  | $x_{4(1)}$  | $x_{5(1)}$  | $x_{6(1)}$  |
| 2         | $y_1$  | $x_{1(2)}$  | $x_{2(2)}$  | $x_{3(2)}$  | $x_{4(2)}$  | $x_{5(2)}$  | $x_{6(2)}$  |
| ⋮         | ⋮      | ⋮           | ⋮           | ⋮           | ⋮           | ⋮           | ⋮           |
| 34        | $y_1$  | $x_{1(34)}$ | $x_{2(34)}$ | $x_{3(34)}$ | $x_{4(34)}$ | $x_{5(34)}$ | $x_{6(34)}$ |

### 3.4 Definisi Operasional

Definisi operasional variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Tingkat Pengangguran Terbuka

TPT adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja.

$$\text{TPT} = \frac{\text{jumlah pengangguran terbuka}}{\text{jumlah angkatan kerja}} 100\%$$

2. Tingkat Partisipasi Angkatan Kerja

Rasio antara angkatan kerja dengan jumlah penduduk usia kerja. Indikator ini dapat dituliskan dengan rumus di bawah ini:

$$\text{TPAK} = \frac{\text{jumlah angkatan kerja}}{\text{jumlah penduduk berusia 15 tahun ke atas}} 100$$

3. Upah Minimum Provinsi (UMP)

Upah Minimum Provinsi (UMP) adalah besaran upah yang dibayarkan pada tiap-tiap wilayah berdasarkan peraturan daerah yang ditetapkan oleh masing-masing wilayah. UMP merupakan suatu standar minimum di suatu provinsi yang digunakan oleh para pengusaha atau pelaku bisnis untuk memberikan upah kepada pegawai karyawan atau buruh didalam lingkungan kerjanya (peraturan Menteri Tenaga Kerja dan Transmigrasi no. 7 2015)

4. Persentase Penduduk 15+ yang lulus SMA/SMK

Persentase penduduk 15 Tahun ke atas yang lulus SMA/SMK adalah jumlah penduduk 15 tahun ke atas yang lulus SMA/SMK dibagi dengan jumlah

penduduk usia 15 tahun ke atas berdasarkan hasil Survei Sosial Ekonomi Nasional.

$$\text{Persentase pddk 15+ lulus SMA/SMK} = \frac{\text{jml pddk usia 15+ yang lulus SMA/SMK}}{\text{jml pddkk usia 15 tahun keatas}}$$

5. Kepadatan Penduduk

Banyaknya penduduk per kilometer persegi. Jumlah penduduk dibagi luas dalam suatu wilayah.

6. Investasi

Banyaknya realisasi penanaman modal dalam negeri dan penanaman modal asing maupun pada setiap provinsi yang bersumber dari data BKPM dan BPS.

7. Laju Pertumbuhan Ekonomi

Laju pertumbuhan ekonomi merupakan salah satu indikator penting untuk mengetahui kondisi ekonomi suatu wilayah dalam periode tertentu. Indikator tersebut berupa perubahan nilai tambah barang dan jasa yang dihitung menggunakan harga yang berlaku pada satu tahun tertentu (tahun dasar).

### 3.5 Tahapan Penelitian

Tahapan yang dilakukan untuk mendapatkan tujuan dalam penelitian ini adalah sebagai berikut.

- I. Mendapatkan interval konfidensi untuk parameter model regresi semiparametrik campuran Spline *Truncated* dan Kernel.

Tahapan untuk mendapatkan tujuan pertama ini adalah sebagai berikut :

- a) Diberikan respon  $y_i$  dengan variabel prediktor komponen parametrik  $x_i$ , sedangkan variabel prediktor  $z_i$  dan  $t_i$  merupakan komponen nonparametrik. Hubungan antara  $y_i$ ,  $x_i$ ,  $z_i$  dan  $t_i$  diasumsikan mengikuti model regresi semiparametrik yang bersifat additif.

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

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



b) Menghampiri kurva-kurva regresi tersebut dengan fungsi-fungsi sebagai berikut :

i. Menghampiri  $f_j(x_{pi})$  dengan menggunakan fungsi linier pada persamaan (2.6)

ii. Menghampiri fungsi  $g_s(t_{qi})$  menggunakan fungsi Spline *Truncated* pada persamaan (2.10)

iii. Menghampiri fungsi  $h_k(z_{ri})$  menggunakan fungsi Kernel pada persamaan (2.16).

c) Tuliskan model regresi semiparametrik campuran spline truncated dan kernel dalam bentuk matrik.

$$\tilde{\mathbf{y}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{G}(k)\tilde{\boldsymbol{\theta}} + \mathbf{D}(\alpha)\tilde{\mathbf{y}} + \tilde{\boldsymbol{\varepsilon}}, \quad \tilde{\boldsymbol{\varepsilon}} \sim IIDN(0, \boldsymbol{\sigma}^2 \mathbf{I})$$

dimana,  $\tilde{f}(x) = \mathbf{X}\tilde{\boldsymbol{\beta}}$ ,  $\tilde{g}(z) = \mathbf{G}(k)\tilde{\boldsymbol{\theta}}$ , dan  $\tilde{h}(t) = \mathbf{D}(\alpha)\tilde{\mathbf{y}}$

d) Membangun komponen *goodness of fit* :

$$\tilde{\boldsymbol{\varepsilon}}' \tilde{\boldsymbol{\varepsilon}} = (\tilde{\mathbf{y}} - \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{G}(k)\tilde{\boldsymbol{\theta}} - \mathbf{D}(\alpha)\tilde{\mathbf{y}})' (\tilde{\mathbf{y}} - \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{G}(k)\tilde{\boldsymbol{\theta}} - \mathbf{D}(\alpha)\tilde{\mathbf{y}})$$

e) Menyelesaikan optimasi OLS :

$$Q(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}) = \underset{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}}{\text{Min}} \|\tilde{\mathbf{y}} - \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{G}(k)\tilde{\boldsymbol{\theta}} - \mathbf{D}(\alpha)\tilde{\mathbf{y}}\|^2$$

dengan menggunakan derivatif parsial sama dengan 0.

f) Menyamakan derivatif parsial tersebut sama dengan nol sehingga didapatkan :  $\hat{\tilde{\boldsymbol{\beta}}}$  dan  $\hat{\tilde{\boldsymbol{\theta}}}$

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

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

g) Mencari distribusi dari  $\hat{\tilde{\boldsymbol{\beta}}}$  dan  $\hat{\tilde{\boldsymbol{\theta}}}$  dengan sebelumnya membentuk vektor parameter  $\tilde{\boldsymbol{\gamma}}$  yang memuat  $\tilde{\boldsymbol{\beta}}$  dan  $\tilde{\boldsymbol{\theta}}$ .

h) Mencari *Pivotal Quantity* untuk parameter  $\tilde{\boldsymbol{\gamma}}$

i) Misalkan  $T_j(x_1, x_2, \dots, x_n, y)$  adalah *Pivotal Quantity* untuk parameter  $\tilde{\boldsymbol{\gamma}}_j$  dan dengan  $j = 1, 2, \dots, (1 + p + q(1 + r))$  Menyelesaikan persamaan dalam probabilitas

$$P(u \leq T_j(x_1, x_2, \dots, x_n) \leq v) = 1 - \alpha$$

j) Menghitung panjang interval konfidensi  $1-\alpha$  yaitu :  $\ell(u, v)$

k) Membentuk fungsi Lagrange :

$$F(u, v, \lambda) = \ell(u, v) + \lambda(\Phi(b) - \Phi(a)) - (1 - \alpha)$$

dimana :

$\Phi$  : fungsi distribusi kumulatif  $N(0,1)$ .

$\lambda$  : konstanta Lagrange.

l) Menghitung derivatif parsial :

$$\frac{dF(u, v, \lambda)}{du} = 0 \quad \frac{dF(u, v, \lambda)}{dv} = 0 \quad \frac{dF(u, v, \lambda)}{d\lambda} = 0$$

m) Menentukan nilai  $(u, v, \lambda)$

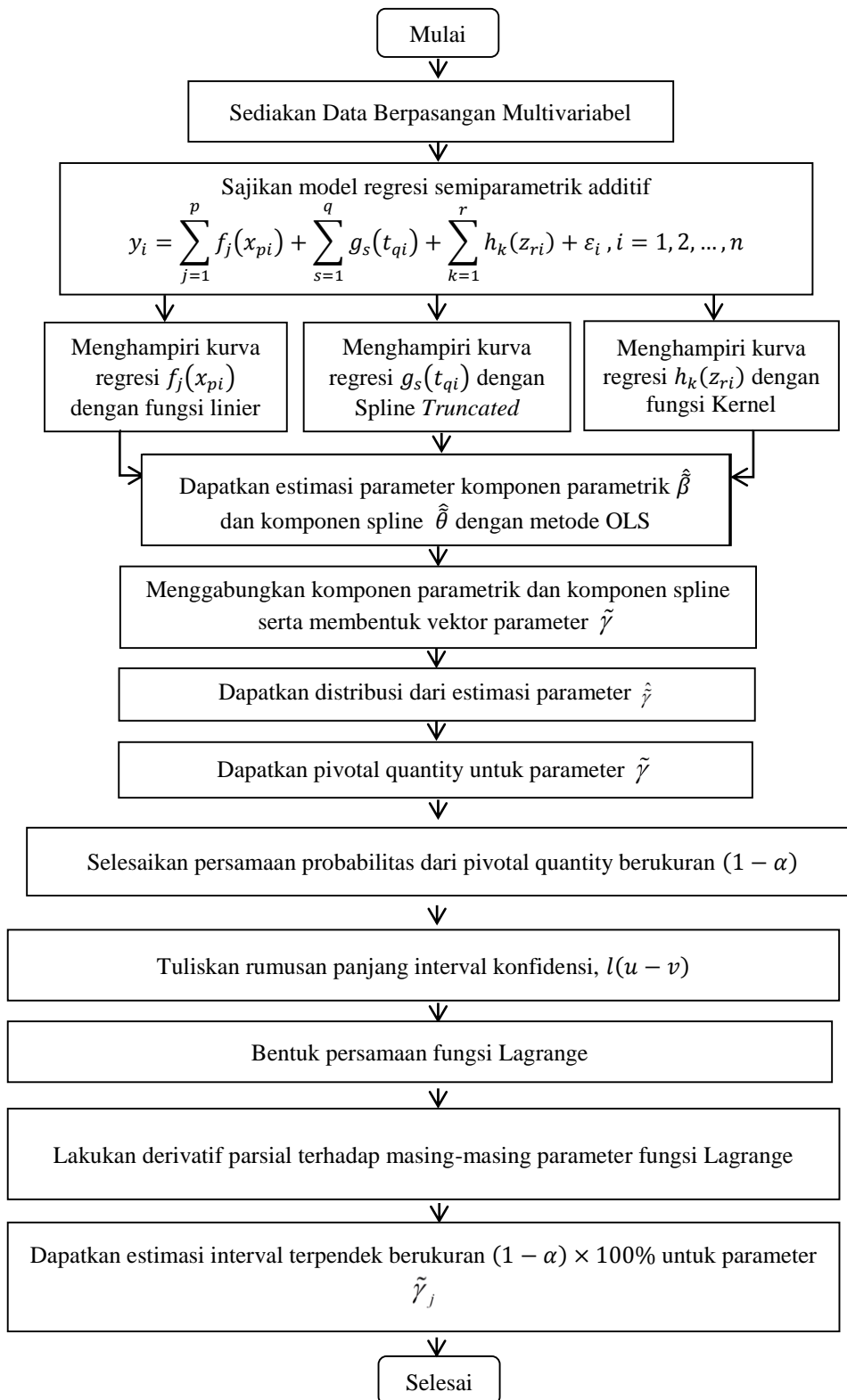
Menghitung interval konfidensi terpendek untuk  $\beta_j$  dan  $\theta_{sk}$  dengan cara menyelesaikan persamaan :

$$P(u \leq T_j(x_1, x_2, \dots, x_n) \leq v) = 1 - \alpha \cdot$$

n) Dari langkah (n) didapatkan :

$$P(u \leq T_j(x_1, x_2, \dots, x_n) \leq v) CI(1 - \alpha) \times 100\% \text{ untuk parameter } \tilde{\gamma}_j \text{ dan untuk.}$$

Diagram alir kajian teoritis untuk menyelesaikan tujuan pertama adalah sebagai berikut.



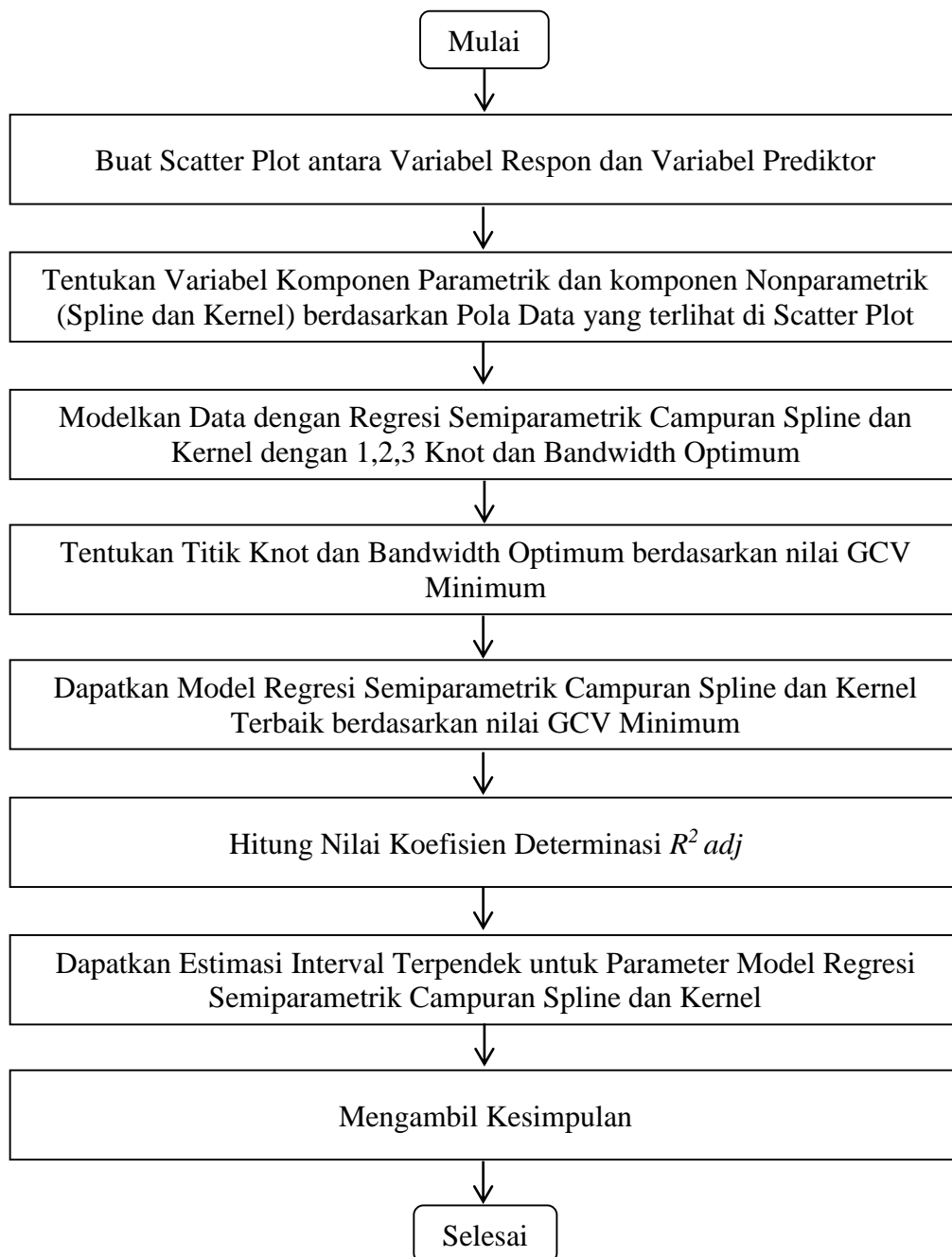
Gambar 3.1 Tahapan Analisis untuk Mendapatkan Estimasi Interval untuk Parameter Komponen Parametrik dan Komponen Spline dalam Regresi Semiparametrik Campuran Spline dan Kernel.

II. Mengaplikasikan estimasi interval parameter pada model regresi semiparametrik campuran spline dan kernel terhadap data Tingkat Pengangguran Terbuka di Indonesia.

Untuk menyelesaikan tujuan kedua, dilakukan langkah-langkah sebagai berikut:

1. Membuat *scatter plot* antara variabel respon dengan variabel prediktor untuk mengetahui pola hubungannya.
2. Menentukan variabel prediktor komponen parametrik, spline dan kernel berdasarkan pola data antara variabel respon dan prediktor.
3. Memodelkan data menggunakan regresi semiparametrik campuran spline dan kernel dengan berbagai titik knot (satu knot, dua knot, tiga knot dan kombinasi knot).
4. Memilih titik knot dan parameter *bandwith* optimal dengan metode GCV.
5. Menetapkan model terbaik dari nilai GCV terkecil.
6. Menghitung  $R^2$  adj sebagai bagian dari kriteria kebaikan model.
7. Menghitung estimasi interval terpendek untuk parameter model regresi semiparametrik campuran spline dan kernel.
8. Mengambil kesimpulan dari model yang terbentuk dan menentukan variabel yang berpengaruh signifikan terhadap variabel respon..

Diagram alir analisis estimasi interval untuk parameter pada data TPT disajikan sebagai berikut.



Gambar 3.2 Tahapan Analisis Estimasi Interval pada Aplikasi data TPT

## BAB 4

### HASIL DAN PEMBAHASAN

Pada sub bab ini akan dibahas tentang estimasi interval untuk parameter komponen parametrik dan komponen spline pada regresi semiparametrik campuran spline dan kernel. Kemudian memodelkan data TPT di Indonesia dengan menggunakan model regresi semiparametrik campuran spline dan kernel dengan titik knot satu knot, dua knot, tiga knot, dan kombinasi knot. Model terbaik ditentukan dengan menggunakan nilai GCV minimum. Selanjutnya diterapkan estimasi interval terpendek parameter komponen parametrik dan komponen spline pada regresi semiparametrik campuran spline dan kernel pada data TPT di Indonesia.

#### 4.1 Model Regresi Semiparametrik Campuran Spline dan Kernel

Diberikan data berpasangan  $(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}, z_{1i}, \dots, z_{ri}, y_i)$ ,  $i = 1, 2, \dots, n$  yang mengikuti model semiparametrik campuran spline dan kernel, dirumuskan sebagai berikut :

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

dengan kurva regresi  $\mu$  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,  $f_j(x_{ji})$  adalah komponen parametrik,  $g_s(t_{si})$  adalah komponen nonparametrik spline,  $h_k(z_{ki})$  adalah komponen nonparametrik kernel, dan  $\varepsilon_i$  adalah error random yang diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varian  $\sigma^2$ .

Kurva regresi  $f_j(x_{ji})$  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 dapat dituliskan dalam bentuk matrik

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

Selanjutnya dalam notasi matriks dapat ditulis menjadi :

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

dimana :

$$\tilde{f}(x) = \begin{bmatrix} f_j(x_{j1}) \\ f_j(x_{j2}) \\ \vdots \\ f_j(x_{jn}) \end{bmatrix}_{n \times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{n \times (p+1)} \quad \text{dan} \quad \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1) \times 1}$$

Kurva regresi  $g_s(t_{si})$  merupakan komponen nonparametrik, dihampiri 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 :

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

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 :

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

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

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

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

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

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

$$\hat{h}_1(z_1) = \mathbf{D}(\alpha_1) \tilde{y}$$

dimana :

$$\hat{h}_1(z_1) = \begin{bmatrix} \hat{h}_1(z_{11}) \\ \hat{h}_1(z_{12}) \\ \vdots \\ \hat{h}_1(z_{1n}) \end{bmatrix}, \tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{D}(\alpha_1) = \begin{bmatrix} n^{-1}W_{\alpha_1}(Z_{11}) & n^{-1}W_{\alpha_2}(Z_{11}) & \dots & n^{-1}W_{\alpha_n}(Z_{11}) \\ n^{-1}W_{\alpha_1}(Z_{12}) & n^{-1}W_{\alpha_2}(Z_{12}) & \dots & n^{-1}W_{\alpha_n}(Z_{12}) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_{\alpha_1}(Z_{1n}) & n^{-1}W_{\alpha_2}(Z_{1n}) & \dots & n^{-1}W_{\alpha_n}(Z_{1n}) \end{bmatrix}$$

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

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

Berdasarkan persamaan (4.3), (4.5) dan (4.8) maka persamaan (4.1) akan menjadi

$$y_i = \sum_{j=0}^p \beta_j x_{ji} + \sum_{s=1}^q \left( \sum_{j=1}^m \theta_{js} t_{si} + \sum_{k=1}^r \theta_{(m+k)s} (t_{si} - K_{ks})_+^m \right) + \sum_{k=1}^r \left( n^{-1} \sum_{i=1}^n W_{\alpha_i}(Z_i) y_i \right) + \varepsilon_i \quad (4.10)$$

Persamaan (4.10) dapat dituliskan dalam notasi matrik sebagai berikut :

$$\tilde{y} = \mathbf{X}\tilde{\beta} + \mathbf{G}(\tilde{k})\tilde{\theta} + \mathbf{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon} \quad (4.11)$$

#### 4.1.1 Estimasi Kurva Regresi Semiparametrik Campuran Spline dan Kernel

Estimasi titik untuk kurva semiparametrik campuran Spline dan Kernel dapat diperoleh dengan menggunakan metode *Ordinary Least Square* (OLS). Dari model persamaan (4.11) diatas bisa dijabarkan sebagai berikut.

$$\begin{aligned} \tilde{\varepsilon} &= \tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} - \mathbf{D}(\tilde{\alpha})\tilde{y} \\ &= (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}\tilde{\beta} - \mathbf{G}(\tilde{k})\tilde{\theta} \end{aligned} \quad (4.12)$$

Selanjutnya akan dicari jumlah kuadrat error yang diberikan oleh :

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

jika disederhanakan menjadi :

$$\begin{aligned}
\tilde{\varepsilon}^T \tilde{\varepsilon} &= \tilde{y}^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y}-2\tilde{\beta}^T\mathbf{X}^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y}-2\tilde{\theta}^T\mathbf{G}(\tilde{k})^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y}+ \\
&\quad 2\tilde{\beta}^T\mathbf{X}^T\mathbf{G}(\tilde{k})\tilde{\theta}+\tilde{\beta}^T\mathbf{X}^T\mathbf{X}\tilde{\beta}+\tilde{\theta}^T\mathbf{G}(\tilde{k})^T\mathbf{G}(\tilde{k})\tilde{\theta} \\
&= Q(\tilde{\beta}, \tilde{\theta}) \tag{4.13}
\end{aligned}$$

Untuk mendapatkan estimasi dari parameter, dilakukan derivatif parsial terhadap  $\tilde{\beta}$  dan  $\tilde{\theta}$ . Derivatif parsial  $Q(\tilde{\beta}, \tilde{\theta})$  diberikan oleh :

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\beta}} = 0 - 2\mathbf{X}^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y} + 2\mathbf{X}^T\mathbf{G}(\tilde{k})\tilde{\theta} + 2\mathbf{X}^T\mathbf{X}\tilde{\beta} \tag{4.14}$$

$$\frac{\partial Q(\tilde{\beta}, \tilde{\theta})}{\partial \tilde{\theta}} = 0 - 2\mathbf{G}(\tilde{k})^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y} + 2\mathbf{G}(\tilde{k})^T\mathbf{X}\tilde{\beta} + 2\mathbf{G}(\tilde{k})^T\mathbf{G}(\tilde{k})\tilde{\theta} \tag{4.15}$$

Jika derivatif parsial (4.14) diatas disamakan dengan nol, sehingga diperoleh persamaan :

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

Sehingga estimator parameter  $\hat{\beta}$  adalah sebagai berikut.

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1} \left( \mathbf{X}^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{X}^T\mathbf{G}(\tilde{k})\hat{\theta} \right) \tag{4.16}$$

Selanjutnya untuk mendapatkan estimator  $\hat{\theta}$  derivatif parsial (4.15) disamakan dengan nol, sehingga diperoleh persamaan :

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

Sehingga estimator parameter  $\hat{\theta}$  adalah sebagai berikut.

$$\hat{\theta} = (\mathbf{G}(\tilde{k})^T\mathbf{G}(\tilde{k}))^{-1} \left( \mathbf{G}(\tilde{k})^T(\mathbf{I}-\mathbf{D}(\tilde{\alpha}))\tilde{y} - \mathbf{G}(\tilde{k})^T\mathbf{X}\hat{\beta} \right) \tag{4.17}$$

Karena estimator parameter  $\tilde{\beta}$  masih mengandung  $\tilde{\theta}$ , dan estimator  $\tilde{\theta}$  masih mengandung  $\tilde{\beta}$  maka dilakukan substitusi antara persamaan (4.16) dengan persamaan (4.17) sebagai berikut.

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

Kemudian komponen yang mengandung  $\hat{\beta}$  dijadikan satu sebagai berikut.

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

Sehingga estimator  $\hat{\beta}$  diperoleh persamaan sebagai berikut.

$$\begin{aligned}
\hat{\beta} &= \left[ \mathbf{I} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} \right]^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \\
&\quad \left[ \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) - \mathbf{X}^T \mathbf{G}(\tilde{k}) (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \right] \tilde{y} \\
&= \mathbf{A}(\tilde{k}, \tilde{\alpha}) \tilde{y} \tag{4.18}
\end{aligned}$$

Dimana  $\mathbf{I}$  adalah matrik identitas,  $\tilde{k}$  adalah titik knot dari  $1, 2, \dots, r$ ,  $\tilde{\alpha}$  adalah besar bandwidth dan matrik

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

Begitu juga untuk estimator parameter  $\tilde{\theta}$ , sehingga diperoleh :

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

Kemudian komponen yang mengandung  $\hat{\theta}$  dijadikan satu di ruas kiri sebagai berikut.

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

Sehingga estimator  $\hat{\theta}$  diperoleh persamaan sebagai berikut :

$$\begin{aligned}
\hat{\theta} &= \left[ \mathbf{I} - (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G}(\tilde{k}) \right]^{-1} (\mathbf{G}(\tilde{k})^T \mathbf{G}(\tilde{k}))^{-1} \\
&\quad \left[ \mathbf{G}(\tilde{k})^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} - \mathbf{G}(\tilde{k})^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{D}(\tilde{\alpha})) \tilde{y} \right] \tilde{y} \\
&= \mathbf{B}(\tilde{k}, \tilde{\alpha}) \tilde{y} \tag{4.19}
\end{aligned}$$

Dimana  $\mathbf{I}$  adalah matrik identitas,  $\tilde{k}$  adalah titik knot dari  $1, 2, \dots, r$ ,  $\tilde{\alpha}$  adalah besar bandwith dan matrik

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

Setelah diperoleh estimator untuk komponen parametrik dan nonparametrik, selanjutnya menentukan estimator gabungan dari model regresi semiparametrik campuran spline dan kernel.

$$\begin{aligned}
\hat{y} &= \hat{\mu}(x, t, z) \\
&= \hat{f}(x) + \hat{g}(t) + \hat{h}(z) \\
&= \mathbf{X}\hat{\beta} + \mathbf{G}(\tilde{k})\hat{\theta} + \mathbf{D}(\tilde{\alpha})\tilde{y} \\
&= \mathbf{XA}(\tilde{k}, \tilde{\alpha})\tilde{y} + \mathbf{GB}(\tilde{k}, \tilde{\alpha})\tilde{y} + \mathbf{D}(\tilde{\alpha})\tilde{y} \\
&= (\mathbf{XA}(\tilde{k}, \tilde{\alpha}) + \mathbf{GB}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha}))\tilde{y} \\
&= \mathbf{M}(\mathbf{X}, \mathbf{G}, \mathbf{D}, \tilde{k}, \tilde{\alpha})\tilde{y} \tag{4.20}
\end{aligned}$$

dimana :

$$\mathbf{M}(\tilde{k}, \tilde{\alpha}) = \mathbf{XA}(\tilde{k}, \tilde{\alpha}) + \mathbf{GB}(\tilde{k}, \tilde{\alpha}) + \mathbf{D}(\tilde{\alpha}) \tag{4.21}$$

#### 4.1.2 Estimasi Interval Untuk Parameter Regresi Semiparametrik Campuran Spline dan Kernel

Estimasi interval merupakan salah satu bagian terpenting dari inferensi statistik. Pada umumnya, untuk mendapatkan estimasi interval digunakan pendekatan Bayesian (Wahba, 1983 ; 1990 dan Wang, 1998). Dalam penelitian ini akan digunakan pendekatan Pivotal Quantity dikarenakan penghitungan matematisnya yang mudah dan tidak diperlukannya distribusi dari prior seperti halnya pendekatan Bayesian. Pivotal Quantity adalah sebuah statistik yang distribusinya tidak memuat parameter. 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.

Dari model regresi semiparametrik campuran spline *truncated* dan kernel seperti pada persamaan (4.10)

$$y_i = \sum_{j=0}^p \beta_j x_{ji} + \sum_{s=1}^q \left( \sum_{j=1}^m \theta_{js} t_{si} + \sum_{k=1}^r \theta_{(m+k)s} (t_{si} - K_{ks})_+^m \right) + \sum_{k=1}^r \left( n^{-1} \sum_{i=1}^n W_{ai}(Z_i) y_i \right) + \varepsilon_i$$

bisa dituliskan dalam bentuk matrik seperti pada persamaan (4.11) sebagai berikut.

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

Dari persamaan diatas akan digabungkan komponen parametrik dan komponen spline yang masing-masing mengandung parameter  $\tilde{\beta}$  dan  $\tilde{\theta}$  dalam bentuk sebagai berikut.

$$\tilde{y} = \mathbf{W}\tilde{\gamma} + \mathbf{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon} \quad (4.22)$$

dengan

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}; \quad \mathbf{W} = [\mathbf{X} : \mathbf{G}]_{n \times ((p+1)+(1+r)q)}; \quad \tilde{\gamma} = \begin{bmatrix} \tilde{\beta} \\ \vdots \\ \tilde{\theta} \end{bmatrix}_{((p+1)+(1+r)q) \times 1}; \quad \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{n \times (1+p)}; \quad \mathbf{G}(k) = \begin{bmatrix} t_{11} (t_{11} - K_{11})_+ \cdots (t_{11} - K_{1r})_+ \cdots t_{q1} (t_{q1} - K_{q1})_+ \cdots (t_{q1} - K_{qr})_+ \\ t_{12} (t_{12} - K_{11})_+ \cdots (t_{12} - K_{1r})_+ \cdots t_{q2} (t_{q2} - K_{q1})_+ \cdots (t_{q2} - K_{qr})_+ \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ t_{ln} (t_{ln} - K_{11})_+ \cdots (t_{ln} - K_{1r})_+ \cdots t_{qn} (t_{qn} - K_{q1})_+ \cdots (t_{qn} - K_{qr})_+ \end{bmatrix}_{n \times (1+r)q}$$

Dimana respon  $\tilde{y}$  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  $n \times (1+p)$  dan  $\mathbf{G}$  adalah matriks yang memuat komponen nonparametrik spline *truncated* berukuran  $((1+r)q) \times 1$ ,  $\tilde{\gamma}$  merupakan vektor parameter dengan ukuran  $((1+p)+(1+r)q) \times 1$  dan  $\tilde{\varepsilon}$  adalah vektor *error* berukuran  $n \times 1$ .

Estimator parameter dari model regresi semiparametrik campuran spline *truncated* dan kernel dapat diperoleh dengan menggunakan metode *Ordinary Least Square* (OLS) dari persamaan (4.22) diatas yang menghasilkan persamaan :

$$\begin{aligned} \tilde{y} &= \mathbf{W}\tilde{\gamma} + \mathbf{D}(\tilde{\alpha})\tilde{y} + \tilde{\varepsilon} \\ (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))\tilde{y} &= \mathbf{W}\tilde{\gamma} + \tilde{\varepsilon} \\ \tilde{y} &= (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^{-1} \mathbf{W}\tilde{\gamma} + (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^{-1} \tilde{\varepsilon} \\ &= \mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^* \end{aligned} \quad (4.23)$$

dimana  $\tilde{\varepsilon}^* = (\mathbf{I} - \mathbf{D}(\tilde{\alpha}))^{-1} \tilde{\varepsilon}$

Dari persamaan (4.23) diatas dapat diperoleh persamaan sebagai berikut.

$$\begin{aligned}\tilde{y} &= \mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^* \\ \tilde{\varepsilon}^* &= \tilde{y} - \mathbf{T}\tilde{\gamma}\end{aligned}\quad (4.24)$$

Selanjutnya dengan menjumlahkan kuadrat error bisa diperoleh hasil sebagai berikut :

$$\begin{aligned}\sum_{i=1}^n \varepsilon_i^{*2} &= \tilde{\varepsilon}^{*T} \tilde{\varepsilon}^* = Q(\tilde{\gamma}) \\ Q(\tilde{\gamma}) &= [\tilde{y} - \mathbf{T}\tilde{\gamma}]^T [\tilde{y} - \mathbf{T}\tilde{\gamma}] \\ &= [\tilde{y}^T - \tilde{\gamma}^T \mathbf{T}^T] [\tilde{y} - \mathbf{T}\tilde{\gamma}] \\ &= \tilde{y}^T \tilde{y} - \tilde{y}^T \mathbf{T} \tilde{\gamma} - \tilde{\gamma}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{T} \tilde{\gamma} \\ &= \tilde{y}^T \tilde{y} - 2\tilde{\gamma}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}\end{aligned}\quad (4.25)$$

Untuk mendapatkan estimator dari parameter, dilakukan derivatif parsial dari persamaan (4.25)  $Q(\tilde{\gamma})$  terhadap  $\tilde{\gamma}$  sebagai berikut.

$$\begin{aligned}\frac{\partial Q(\tilde{\gamma})}{\partial \tilde{\gamma}} &= \frac{\tilde{y}^T \tilde{y} - 2\tilde{\gamma}^T \mathbf{T}^T \tilde{y} + \tilde{\gamma}^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}}{\partial \tilde{\gamma}} \\ &= -2\mathbf{T}^T \tilde{y} + 2\mathbf{T}^T \mathbf{T} \tilde{\gamma}\end{aligned}\quad (4.26)$$

Apabila persamaan (4.26) disamakan dengan nol, maka akan diperoleh persamaan berikut.

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

selanjutnya persamaan dapat ditulis dalam bentuk :

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

Sehingga diperoleh estimator parameter sebagai berikut.

$$\hat{\gamma} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{y}\quad (4.27)$$

Persamaan (4.27) dapat diringkas menjadi bentuk persamaan :

$$\hat{\gamma} = \mathbf{F}(\tilde{k}, \tilde{\alpha}) \tilde{y}\quad (4.28)$$

dimana :

$$\mathbf{F}(\tilde{k}, \tilde{\alpha}) = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T\quad (4.29)$$

Setelah didapatkan estimator dari  $\tilde{\gamma}$ , langkah selanjutnya akan dicari distribusi dari  $\hat{\gamma}$  dengan tahapan sebagai berikut. Apabila diasumsikan  $\tilde{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ , 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.

$$\begin{aligned} E(\tilde{y}) &= E(\mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^*) \\ &= \mathbf{T}\tilde{\gamma} + 0 \\ &= \mathbf{T}\tilde{\gamma} \end{aligned} \tag{4.30}$$

$$\begin{aligned} Var(\tilde{y}) &= Var(\mathbf{T}\tilde{\gamma} + \tilde{\varepsilon}^*) \\ &= 0 + Var(\tilde{\varepsilon}^*) \\ &= Var(\tilde{\varepsilon}^*) \\ &= \sigma^2 \mathbf{I} \end{aligned} \tag{4.31}$$

Berdasarkan (4.30) dan (4.31) 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  $\hat{\gamma}$ .

$$\begin{aligned} E(\hat{\gamma}) &= E(\mathbf{F}(\tilde{k}, \tilde{\alpha})\tilde{y}) \\ &= \mathbf{F}(\tilde{k}, \tilde{\alpha})E(\tilde{y}) \\ &= (\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r E(\tilde{y}) \\ &= (\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \mathbf{T} \tilde{\gamma} \\ &= \tilde{\gamma} \end{aligned} \tag{4.32}$$

$$\begin{aligned} Var(\hat{\gamma}) &= Var(\mathbf{F}(\tilde{k}, \tilde{\alpha})\tilde{y}) \\ &= \mathbf{F}(\tilde{k}, \tilde{\alpha})Var(\tilde{y})\mathbf{F}(\tilde{k}, \tilde{\alpha})^T \\ &= ((\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \sigma^2 \mathbf{I} ((\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r)^T \\ &= \sigma^2 \mathbf{I} (\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \mathbf{T} (\mathbf{T}^r \mathbf{T})^{-1} \\ &= \sigma^2 \mathbf{I} (\mathbf{T}^r \mathbf{T})^{-1} \end{aligned} \tag{4.33}$$

$$\hat{\gamma} \sim N(\tilde{\gamma}, \sigma^2 (\mathbf{T}^r \mathbf{T})^{-1}) \tag{4.34}$$

Setelah didapatkan estimasi dan distribusi dari  $\hat{\gamma}$  seperti yang telah dijabarkan pada persamaan (4.27) hingga persamaan (4.34). Selanjutnya akan dicari estimasi interval untuk parameter dengan variansi  $\sigma^2$  diketahui dan variansi  $\sigma^2$  tidak diketahui.



### 4.1.3 Estimasi Interval untuk $\sigma^2$ Diketahui

Pada subbab ini akan dirancang estimasi interval terpendek  $(1-\alpha)\times 100\%$  untuk  $\tilde{\gamma}_j$  dimana  $\sigma^2$  diketahui. Diambil sebuah transformasi :

$$Z_j(x_1, x_2, \dots, x_n) = \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 w_{jj}}} \quad (4.35)$$

dengan  $w_{jj}$  merupakan elemen diagonal ke- $jj$  dari  $(\mathbf{T}^r \mathbf{T})^{-1}$  dan variabel random  $Z_j \sim N(0,1)$ . Tahapan selanjutnya akan dicari ekspektasi dan varians dari  $Z_j$  untuk mengetahui distribusi dan sifat-sifat dari  $Z_j$ . Ekspektasi dan varians dari  $Z_j$  diberikan sebagai berikut.

$$\begin{aligned} E(Z_j) &= \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 w_{jj}}} \\ &= \frac{1}{\sqrt{\sigma^2 w_{jj}}} E(\hat{\gamma}_j - \tilde{\gamma}_j) \\ &= \frac{1}{\sqrt{\sigma^2 w_{jj}}} E(\tilde{\gamma}_j - \tilde{\gamma}_j) \\ &= 0 \\ \text{Var}(Z_j) &= \text{Var}\left(\frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 w_{jj}}}\right) \\ &= \left(\frac{1}{\sqrt{\sigma^2 w_{jj}}}\right)^2 \text{Var}(\hat{\gamma}_j - \tilde{\gamma}_j) \\ &= \left(\frac{1}{\sqrt{\sigma^2 w_{jj}}}\right)^2 \sigma^2 w_{jj} \\ &= 1 \end{aligned}$$

Akibatnya, karena  $Z_j$  merupakan kombinasi linier dari  $\hat{\gamma}$  dan  $\hat{\gamma} \sim N(\tilde{\gamma}, \sigma(\mathbf{T}^r \mathbf{T})^{-1})$  maka  $Z_j \sim N(0,1)$ . Karena  $\sigma^2$  diketahui maka  $Z_j(x_1, x_2, \dots, x_n)$  merupakan *pivotal quantity* untuk parameter  $\tilde{\gamma}$ . Estimasi interval  $(1-\alpha)$  dapat diperoleh dengan menyelesaikan persamaan dalam probabilitas

$$P(u \leq Z_j(x_1, x_2, \dots, x_n) \leq v) = 1 - \alpha \quad (4.36)$$

dengan  $u$  dan  $v$  merupakan elemen bilangan riil, dimana  $u < v$ . Apabila persamaan (4.35) di substitusikan ke dalam persamaan (4.36), maka akan didapatkan

$$P\left(u \leq \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 w_{jj}}} \leq v\right) = 1 - \alpha \quad (4.37)$$

persamaan diatas ekuivalen dengan persamaan (4.38) dan persamaan (4.39), seperti dibawah ini.

$$\tilde{\gamma}_j \leq \hat{\gamma}_j - u\sqrt{\sigma^2 w_{jj}} \quad (4.38)$$

$$\tilde{\gamma}_j \geq \hat{\gamma}_j - v\sqrt{\sigma^2 w_{jj}} \quad (4.39)$$

Berdasarkan kedua persamaan diatas (4.38) dan (4.39) didapatkan estimasi interval untuk parameter regresi semiparametrik campuran spline *truncated* dan kernel adalah sebagai berikut.

$$P\left(\hat{\gamma}_j - v\sqrt{\sigma^2 w_{jj}} \leq \tilde{\gamma}_j \leq \hat{\gamma}_j - u\sqrt{\sigma^2 w_{jj}}\right) = 1 - \alpha \quad (4.40)$$

Pada penelitian ini akan dicari estimasi interval terpendek, sehingga harus ditentukan nilai  $u$  dan  $v$  agar panjang estimasi interval  $\ell(u, v)$  pada persamaan (4.40) memiliki jarak terpendek. Untuk mendapatkan estimasi interval terpendek, dapat dilakukan dengan menyelesaikan optimasi bersyarat sebagai berikut dengan menggunakan metode *lagrange*.

$$\text{Min}_{u,v \in R} \{\ell(u, v)\} = \text{Min}_{u,v \in R} \{(u, v)\sqrt{\sigma^2 w_{jj}}\}$$

Dengan syarat pada persamaan (4.41) yaitu

$$\int_u^v \phi(z) dz = 1 - \alpha, \text{ atau } \Phi(v) - \Phi(u) - (1 - \alpha) = 0 \quad (4.41)$$

Fungsi  $\phi$  merupakan distribusi probabilitas  $N(0,1)$  dan  $\Phi$  merupakan distribusi probabilitas kumulatif  $N(0,1)$ . Kemudian dapat dibentuk fungsi *lagrange*

$$F(u, v, \lambda) = (u - v)\sqrt{\sigma^2 w_{jj}} + \lambda(\Phi(v) - \Phi(u) - (1 - \alpha)) \quad (4.42)$$

dimana  $u, v, \lambda$  merupakan konstanta *lagrange*. Langkah selanjutnya adalah melakukan derivatif parsial pada persamaan (4.42) terhadap  $u, v$  dan  $\lambda$

$$\frac{\partial F(u, v, \lambda)}{\partial u} = 0 \Rightarrow -\sqrt{\sigma^2 w_{jj}} - \lambda \Phi'(u) \quad (4.43)$$

$$\frac{\partial F(u, v, \lambda)}{\partial v} = 0 \Rightarrow \sqrt{\sigma^2 w_{jj}} + \lambda \Phi'(v) = 0 \quad (4.44)$$

$$\frac{\partial F(u, v, \lambda)}{\partial \lambda} = 0 \Rightarrow (\Phi(v) - \Phi(u) - (1 - \alpha)) = 0 \quad (4.45)$$

Selanjutnya dilakukan eliminasi persamaan (4.43) dengan (4.44) dan diperoleh hasil sebagai berikut.

$$\lambda(\Phi'(v) - \Phi'(u)) = 0 \quad (4.46)$$

Persamaan (4.46) dapat juga dituliskan dalam berikut.

$$\phi(v) = \phi(u) \quad (4.47)$$

Hasil perhitungan pada persamaan (4.47) mempunyai 2 (dua) kemungkinan hasil yaitu :

1.  $u = v$

Hasil ini tidak memenuhi karena tidak sesuai dengan persamaan (4.45).

2.  $u = -v$

Kemungkinan kedua, yaitu  $u = -v$  memenuhi.

Dari hasil ini selanjutnya  $u = -v$  disubstitusikan ke dalam persamaan (4.40) maka akan menghasilkan persamaan (4.48) seperti dibawah ini.

$$P\left(\hat{\gamma}_j - v\sqrt{\sigma^2 w_{jj}} \leq \tilde{\gamma}_j \leq \hat{\gamma}_j + v\sqrt{\sigma^2 w_{jj}}\right) = 1 - \alpha \quad (4.48)$$

dimana nilai  $v$  dapat diperoleh dari  $\int_u^v \varphi(t) dz = \frac{\alpha}{2}$  dengan melihat tabel distribusi normal standar  $\sim N(0,1)$ .

#### 4.1.4 Estimasi Interval untuk $\sigma^2$ Tidak Diketahui

Pada subbab ini akan dibahas tentang estimasi interval untuk  $\tilde{\gamma}_j$  dimana  $\sigma^2$  tidak diketahui. Sebelumnya sudah diketahui bahwa persamaan (4.35) adalah *pivotal quantity* untuk  $\tilde{\gamma}_j$  saat  $\sigma^2$  diketahui, namun seringkali dalam aplikasinya sering ditemui kasus dengan  $\sigma^2$  tidak diketahui.

Oleh karena itu  $\sigma^2$  pada persamaan (4.35) akan diduga dengan MSE, sehingga akan didapatkan *pivotal quantity* seperti pada persamaan (4.49).

$$T_j(x_1, x_2, \dots, x_n) = \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{MSE(\mathbf{T}^r \mathbf{T})_{jj}^{-1}}} \quad (4.49)$$

dimana  $(\mathbf{T}^r \mathbf{T})_{jj}^{-1}$  adalah elemen diagonal ke- $jj$  dari matriks  $(\mathbf{T}^r \mathbf{T})^{-1}$  dan MSE nya adalah sebagai berikut.

$$MSE = \frac{\sum_{i=1}^n (\tilde{y}_i - \hat{y}_i)^2}{n - (1 + p + q(r+1))} \quad (4.50)$$

Selanjutnya akan dijabarkan pembilang dari MSE :

$$\begin{aligned} \sum_{i=1}^n (\tilde{y}_i - \hat{y}_i)^2 &= (\tilde{\mathbf{y}} - \hat{\mathbf{y}})^T (\tilde{\mathbf{y}} - \hat{\mathbf{y}}) \\ &= (\tilde{\mathbf{y}} - \mathbf{T} \hat{\boldsymbol{\gamma}})^T (\tilde{\mathbf{y}} - \mathbf{T} \hat{\boldsymbol{\gamma}}) \\ &= (\tilde{\mathbf{y}} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \tilde{\mathbf{y}})^T (\tilde{\mathbf{y}} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \tilde{\mathbf{y}}) \\ &= (\tilde{\mathbf{y}}^r - \tilde{\mathbf{y}}^r \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r)^T (\tilde{\mathbf{y}} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \tilde{\mathbf{y}}) \\ &= \tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r)^T (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}} \\ &= \tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r + \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}} \\ &= \tilde{\mathbf{y}}^r (\mathbf{I} - 2\mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r + \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}} \\ &= \tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}} \\ &= \tilde{\mathbf{y}}^r \mathbf{A} \tilde{\mathbf{y}}, \quad \mathbf{A} = (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \end{aligned} \quad (4.51)$$

Variabel random  $T = \frac{m_1}{\sqrt{\frac{m_2}{a}}} \sim t_{(a)}$  jika dan hanya jika memenuhi ketiga syarat

berikut yaitu :

- i)  $m_1 \sim N(0,1)$
- ii)  $m_2 \sim \chi_{(a)}$  serta
- iii)  $m_1$  dan  $m_2$  saling independen.

Variabel random  $T_j$  pada persamaan (4.49) berdistribusi  $t_{(n-1+p+q(r+1))}$  karena telah memenuhi ketiga syarat tersebut. Untuk mempermudah pembuktiannya, akan dilakukan sedikit penjabaran pada persamaan (4.49) dan diperoleh hasil pada persamaan (4.52).

$$\begin{aligned}
T_j(x_1, x_2, \dots, x_n) &= \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\frac{\tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))} (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}} \\
&= \frac{\frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{(\mathbf{T}^r \mathbf{T})_{jj}^{-1}}}}{\sqrt{\frac{\tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))}}} \\
&= \frac{\frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}}}{\sqrt{\frac{\left( \frac{\tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}}}{\sigma^2} \right)}{n - (1 + p + q(r+1))}}} \\
&= \frac{m_1}{\sqrt{\frac{m_2}{n - (1 + p + q(r+1))}}}
\end{aligned} \tag{4.52}$$

dimana :

$$m_1 = \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}} \text{ dan } m_2 = \left( \frac{\tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}}}{\sigma^2} \right) = \left( \frac{\tilde{\mathbf{y}}^r \mathbf{A} \tilde{\mathbf{y}}}{\sigma^2} \right)$$

Berikut adalah pembuktian bahwa varaiabel random  $T_j$  berdistribusi  $t$  dengan derajat bebas  $n - (1 + p + q(r+1))$ .

i) Membuktikan  $m_1 \sim N(0,1)$

$$m_1 = \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}}, \text{ dikarenakan } m_1 \text{ merupakan kombinasi linier dari } \hat{\gamma}_j \text{ maka}$$

$m_1$  berdistribusi normal dengan mean  $E(m_1)$  dan varians  $Var(m_1)$ .

$$\begin{aligned}
E(m_1) &= \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}} \\
&= \frac{1}{\sqrt{\sigma^2 (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}} E(\hat{\gamma}_j - \tilde{\gamma}_j) \\
&= \frac{1}{\sqrt{\sigma^2 (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}} E(\tilde{\gamma}_j - \tilde{\gamma}_j) \\
&= 0
\end{aligned} \tag{4.53}$$

$$\begin{aligned}
\text{Var}(m_1) &= \text{Var}\left(\frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 (\mathbf{T}^T \mathbf{T})_{jj}^{-1}}}\right) \\
&= \left(\frac{1}{\sqrt{\sigma^2 (\mathbf{T}^T \mathbf{T})_{jj}^{-1}}}\right)^2 \text{Var}(\hat{\gamma}_j - \tilde{\gamma}_j) \\
&= \left(\frac{1}{\sqrt{\sigma^2 (\mathbf{T}^T \mathbf{T})_{jj}^{-1}}}\right)^2 \sigma^2 (\mathbf{T}^T \mathbf{T})_{jj}^{-1} \\
&= 1
\end{aligned} \tag{4.54}$$

Berdasarkan persamaan (4.53) dan (4.54) maka didapatkan  $m_1$  berdistribusi normal dengan mean 0 dan varians 1 (terbukti).

ii) Membuktikan  $m_2 \sim \chi^2_{n-(1+p+q(r+1))}$

$$m_2 = \left(\frac{\tilde{\mathbf{y}}^T \mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \tilde{\mathbf{y}}}{\sigma^2}\right) = \left(\frac{\tilde{\mathbf{y}}^T \mathbf{A} \tilde{\mathbf{y}}}{\sigma^2}\right)$$

Dengan menggunakan Teorema 2.6.4 maka akan diperlihatkan (ii) dengan langkah sebagai berikut.

Akan ditunjukkan matriks  $\mathbf{A} = (\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T)$  adalah simetris dan idempotent, matriks  $\mathbf{A}$  dikatakan simetris jika  $\mathbf{A}^T = \mathbf{A}$ . Berikut ini adalah pembuktian bahwa  $\mathbf{A}$  adalah simetris.

$$\begin{aligned}
\mathbf{A}^T &= (\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T)^T \\
&= \mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \\
&= \mathbf{A}
\end{aligned} \tag{4.55}$$

Dari penjabaran diatas terlihat bahwa matriks  $\mathbf{A}$  simetris. Selanjutnya kan dibuktikan juga bahwa matriks  $\mathbf{A}$  *idempotent*. Matriks  $\mathbf{A}$  dikatakan *idempotent* jika  $\mathbf{A}\mathbf{A} = \mathbf{A}$ . Berikut ini adalah pembuktian bahwa  $\mathbf{A}$  adalah *idempotent*.

$$\begin{aligned}
\mathbf{A}\mathbf{A} &= (\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T)(\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T) \\
&= (\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T + \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T) \\
&= (\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T + \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T) \\
&= (\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T) = \mathbf{A}
\end{aligned} \tag{4.56}$$

Berdasarkan hasil dari persamaan (4.55) dan (4.56) menunjukkan bahwa matriks  $\mathbf{A}$  simetris dan idempotent, sehingga

$$m_2 \sim \chi^2(\text{rank}(\mathbf{A}), \tilde{\boldsymbol{\mu}}^T \mathbf{A} \tilde{\boldsymbol{\mu}} / 2\sigma^2) \tag{4.57}$$

dimana  $\tilde{\mu} = E(\tilde{y}) = \mathbf{T}\tilde{\gamma}$

Selanjutnya akan dicari rank dari matriks  $\mathbf{A}$  yang merupakan matriks simetris dan *idempoten* sehingga  $\text{rank}(\mathbf{A}) = \text{trace}(\mathbf{A})$  berdasarkan Teorema 2.6.3. dengan langkah sebagai berikut :

$$\begin{aligned}
 \text{rank}(\mathbf{A}) &= \text{trace}(\mathbf{I} - \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T) \\
 &= \text{trace}(\mathbf{I}) - \text{trace}(\mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T) \\
 &= n - \text{trace}(\mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T) \\
 &= n - \text{trace}(\mathbf{I}_{(1+p+q(r+1))}) \\
 &= n - (1 + p + q(r+1))
 \end{aligned} \tag{4.58}$$

selanjutnya akan dihitung  $\frac{\tilde{\mu}^T \mathbf{A} \tilde{\mu}}{2\sigma^2}$  sebagai berikut.

$$\begin{aligned}
 \frac{\tilde{\mu}^T \mathbf{A} \tilde{\mu}}{2\sigma^2} &= \frac{(\mathbf{T}\tilde{\gamma})^T (\mathbf{I} - \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T) \mathbf{T}\tilde{\gamma}}{2\sigma^2} \\
 &= \frac{\tilde{\gamma}^T \mathbf{T}^T (\mathbf{I} - \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T) \mathbf{T}\tilde{\gamma}}{2\sigma^2} \\
 &= \frac{\tilde{\gamma}^T \mathbf{T}^T \mathbf{T} \tilde{\gamma} - \tilde{\gamma}^T \mathbf{T}^T \mathbf{T} (\mathbf{T}^T\mathbf{T})^{-1} \mathbf{T}^T \mathbf{T} \tilde{\gamma}}{2\sigma^2} \\
 &= \frac{\tilde{\gamma}^T \mathbf{T}^T \mathbf{T} \tilde{\gamma} - \tilde{\gamma}^T \mathbf{T}^T \mathbf{T} \tilde{\gamma}}{2\sigma^2} \\
 &= 0
 \end{aligned}$$

Dari uraian diatas maka dapat disimpulkan bahwa  $m_2 \sim \chi^2_{(n-(1+p+q(r+1)))}$ .

iii) Membuktikan  $m_1$  dan  $m_2$  *independen* atau saling bebas.

Sesuai dengan Teorema 2.6.5 maka, jika  $\tilde{y} \sim N(\mathbf{T}\tilde{\gamma}, \sigma^2\mathbf{I})$  maka  $\tilde{y}\mathbf{A}\tilde{y}$  dan  $\mathbf{B}\tilde{y}$  akan independen jika dan hanya jika  $\mathbf{BA} = 0$ . Karena  $\hat{\gamma} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\tilde{y}$  maka diperoleh  $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\tilde{y} = \mathbf{B}\tilde{y}$  dengan  $\mathbf{B} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$  dan  $\mathbf{A} = \mathbf{I} - \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$  maka akan dibuktikan  $\mathbf{BA} = 0$ .

$$\begin{aligned}
 \mathbf{BA} &= ((\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T)(\mathbf{I} - \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T) \\
 &= (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T - (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T \\
 &= (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T - (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T \\
 &= 0
 \end{aligned}$$

Sehingga dapat dikatakan bahwa  $m_1$  dan  $m_2$  *independen* atau saling bebas.

Berdasarkan kesimpulan pada poin (i), (ii) dan poin (iii) dapat ditarik kesimpulan bahwa  $T_j$  berdistribusi t dengan derajat bebas  $(n - (1 + p + q(r + 1)))$ . Sehingga  $T_j(x_1, x_2, \dots, x_n)$  merupakan pivotal quantity untuk parameter  $\tilde{\gamma}_j$  saat  $\sigma^2$  tidak diketahui. Estimasi interval  $1 - \alpha$  dapat diperoleh dengan menyelesaikan persamaan dalam probabilitas

$$P(u \leq T_j(x_1, x_2, \dots, x_n) \leq v) = 1 - \alpha \quad (4.59)$$

Dengan  $u$  dan  $v$  merupakan elemen bilangan riil, dimana  $u < v$ . Apabila persamaan (4.49) disubstitusikan ke dalam persamaan (4.59) maka bentuknya akan menjadi

$$P\left(u \leq \frac{\hat{\gamma}_j - \tilde{\gamma}_j}{\sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}}} \leq v\right) = 1 - \alpha$$

persamaan diatas ekuivalen dengan

$$\begin{aligned} \text{i)} \quad & u \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} \leq \hat{\gamma}_j - \tilde{\gamma}_j \\ & \tilde{\gamma}_j \leq \hat{\gamma}_j - u \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} \end{aligned} \quad (4.60)$$

$$\begin{aligned} \text{ii)} \quad & \hat{\gamma}_j - \tilde{\gamma}_j \leq v \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} \\ & \tilde{\gamma}_j \geq \hat{\gamma}_j - v \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} \end{aligned} \quad (4.61)$$

Berdasarkan persamaan (4.60) dan (4.61) didapatkan estimasi interval untuk parameter regresi semiparametrik campuran spline *truncated* dan kernel saat  $\sigma^2$  tidak diketahui adalah sebagai berikut.

$$\begin{aligned} & P\left(\hat{\gamma}_j - v \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} \leq \tilde{\gamma}_j \leq \right. \\ & \left. \hat{\gamma}_j - u \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}}\right) = 1 - \alpha \end{aligned} \quad (4.62)$$



Pada penelitian ini akan dicari estimasi interval terpendek, sehingga harus ditentukan nilai  $u$  dan  $v$  agar panjang estimasi interval  $\ell(u, v)$  pada persamaan (4.62) memiliki jarak terpendek. Untuk mendapatkan estimasi interval terpendek, dapat dilakukan dengan menyelesaikan optimasi bersyarat sebagai berikut dengan menggunakan metode *lagrange*.

$$\text{Min}_{u, v \in R} \{ \ell(u, v) \} = \text{Min}_{u, v \in R} \left\{ (u - v) \sqrt{\frac{\tilde{y}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{y}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} \right\} \quad (4.63)$$

dengan syarat

$$\int_u^v \varphi(t) dz = 1 - \alpha, \text{ atau } \Phi(v) - \Phi(u) - (1 - \alpha) = 0 \quad (4.64)$$

Fungsi  $\varphi$  merupakan distribusi probabilitas  $t_{(n-p(r+1))}$  dan  $\Phi$  merupakan distribusi probabilitas kumulatif  $t_{(n-p+q(r+1))}$ . Kemudian dapat dibentuk fungsi *lagrange* untuk optimasi

$$F(u, v, \lambda) = (v - u) \sqrt{\frac{\tilde{y}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{y}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} + \lambda (\Phi(v) - \Phi(u) - (1 - \alpha)) \quad (4.65)$$

dimana  $u, v, \lambda$  merupakan konstanta *lagrange*. Langkah selanjutnya adalah melakukan derivatif parsial pada persamaan (4.65) terhadap  $u, v$  dan  $\lambda$ .

$$\frac{\partial F(u, v, \lambda)}{\partial u} = 0 \Rightarrow -\sqrt{\frac{\tilde{y}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{y}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} - \lambda \Phi'(u) = 0 \quad (4.66)$$

$$\frac{\partial F(u, v, \lambda)}{\partial v} = 0 \Rightarrow \sqrt{\frac{\tilde{y}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{y}}{n - (1 + p + q(r + 1))} (\mathbf{T}^T \mathbf{T})_{jj}^{-1}} + \lambda \Phi'(v) = 0 \quad (4.67)$$

$$\frac{\partial F(u, v, \lambda)}{\partial \lambda} = 0 \Rightarrow (\Phi(v) - \Phi(u) - (1 - \alpha)) = 0 \quad (4.68)$$

Selanjutnya dilakukan eliminasi persamaan (4.66) dengan (4.67) dan diperoleh hasil sebagai berikut.

$$\lambda (\Phi'(v) - \Phi'(u)) = 0 \quad (4.69)$$

Persamaan (4.69) dapat juga dituliskan dalam berikut.

$$\phi(v) = \phi(u) \quad (4.70)$$

Hasil perhitungan pada persamaan (4.70) mempunyai 2 (dua) kemungkinan hasil yaitu :

1.  $u = v$

Hasil ini tidak memenuhi karena tidak sesuai dengan persamaan (4.68).

2.  $u = -v$

Kemungkinan kedua, yaitu  $u = -v$  memenuhi.

Dari hasil ini selanjutnya  $u = -v$  disubstitusikan ke dalam persamaan (4.62) maka akan menghasilkan persamaan (4.71) seperti dibawah ini.

$$P \left( \hat{\gamma}_j - v \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))}} (\mathbf{T}^T \mathbf{T})_{jj}^{-1} \leq \tilde{\gamma}_j \leq \hat{\gamma}_j + v \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))}} (\mathbf{T}^T \mathbf{T})_{jj}^{-1} \right) = 1 - \alpha \quad (4.71)$$

dimana nilai  $v$  dapat diperoleh dari  $\int_u^v \varphi(t) dz = 1 - \alpha$  dengan melihat nilai pada tabel distribusi  $t_{\frac{\alpha}{2}, (n - (1 + p + q(r+1)))}$ .

Berdasarkan uraian diatas didapatkan hasil yaitu estimasi interval untuk parameter regresi semiparametrik campuran spline *truncated* dan kernel untuk  $\sigma^2$  diketahui dan  $\sigma^2$  tidak diketahui. Adapun estimasi interval saat  $\sigma^2$  diketahui adalah sebagai berikut.

$$P \left( \hat{\gamma}_j - z_{\frac{\alpha}{2}} \sqrt{\sigma^2 w_{jj}} \leq \tilde{\gamma}_j \leq \hat{\gamma}_j + z_{\frac{\alpha}{2}} \sqrt{\sigma^2 w_{jj}} \right) = 1 - \alpha \quad (4.72)$$

dengan  $w_{jj}$  merupakan elemen diagonal ke- $j$  dari  $(\mathbf{T}^T \mathbf{T})^{-1}$ . Estimasi interval saat  $\sigma^2$  tidak diketahui adalah sebagai berikut.

$$P \left( \hat{\gamma}_j - t_{\frac{\alpha}{2}, (n - (1 + p + q(r+1)))} \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))}} (\mathbf{T}^T \mathbf{T})_{jj}^{-1} \leq \tilde{\gamma}_j \leq \hat{\gamma}_j + t_{\frac{\alpha}{2}, (n - (1 + p + q(r+1)))} \sqrt{\frac{\tilde{\mathbf{y}}^T [\mathbf{I} - \mathbf{T}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T] \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))}} (\mathbf{T}^T \mathbf{T})_{jj}^{-1} \right) = 1 - \alpha \quad (4.73)$$

## 4.2 Aplikasi Pada Data Tingkat Pengangguran Terbuka di Indonesia

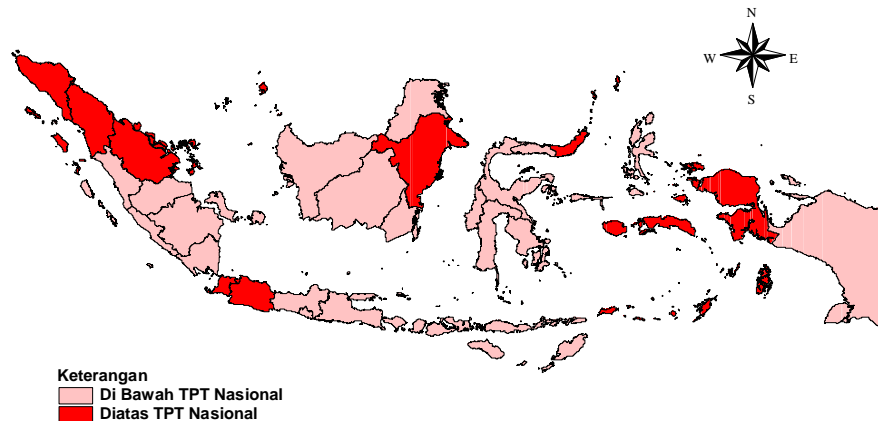
Pada sub bab 4.1 sudah dibahas terkait model dan estimasi interval untuk parameter regresi semiparametrik campuran spline *truncated* dan kernel. Selanjutnya akan diaplikasikan pada data TPT Tahun 2016 di Indonesia. Variabel respon yang digunakan adalah TPT ( $y$ ), sedangkan variabel prediktor yang digunakan sebanyak enam variabel yaitu tingkat partisipasi angkatan kerja ( $x_1$ ), upah minimum provinsi ( $x_2$ ), persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK, kepadatan penduduk ( $x_4$ ), investasi ( $x_5$ ) dan laju pertumbuhan ekonomi ( $x_6$ ). Unit observasi yang digunakan sebanyak 34 provinsi. Berikut ini akan dibahas penerapan dari model regresi semiparametrik campuran spline *truncated* dan kernel pada data TPT di Indonesia tahun 2016.

### 4.2.1 Analisis Deskriptif Data

Tahapan analisis deskriptif diperlukan untuk eksplorasi data sebelum melakukan pemodelan dan estimasi interval untuk parameter, akan dilakukan eksplorasi data melalui analisis deskriptif. Berikut ini disajikan analisis deskriptif pada data yang digunakan..

Beberapa provinsi di Indonesia masih memiliki TPT di atas angka nasional, yaitu sebesar 32,35 persen dari 34 provinsi. Sebagian besar berada di wilayah Indonesia bagian barat. Adapun provinsi yang memiliki TPT di atas angka nasional adalah Provinsi Aceh, Provinsi Sumatra Utara, Provinsi Riau, Provinsi Kepulauan Riau, Provinsi Jawa Barat, Provinsi Daerah Khusus Ibukota (DKI) Jakarta, Provinsi Banten, Provinsi Kalimantan Timur, Provinsi Sulawesi Utara, Provinsi Maluku dan Provinsi Papua Barat.

Gambar 4.1 menunjukkan pengelompokan TPT menurut provinsi yang dibagi ke dalam 2 kelompok dengan menggunakan TPT nasional sebagai cut off point. Lima Provinsi yang memiliki TPT terendah adalah Provinsi Bali, Provinsi Bangka Belitung, Provinsi DIY, Provinsi Sulawesi Tenggara dan Provinsi Gorontalo. Sedangkan lima Provinsi yang memiliki TPT tertinggi adalah Provinsi Banten, Provinsi Jawa Barat, Provinsi Kalimantan Timur, Provinsi Kepulauan Riau dan Provinsi Aceh.



Gambar 4.1 Peta TPT Menurut Provinsi di Indonesia

Karakteristik masing-masing variabel, baik variabel respon maupun variabel prediktor dapat dilihat pada Tabel 4.1. Berdasarkan Tabel 4.1, variabel respon ( $y$ ) yaitu TPT Indonesia pada tahun 2016 memiliki nilai rata-rata sebesar 4,99 tahun dengan standar deviasi sebesar 1,93. Adapun range nilai TPT di Indonesia yang meliputi 34 provinsi adalah 7,03. Nilai TPT tertinggi sebesar 8,92 persen. yaitu di Provinsi Banten. Sedangkan nilai TPT terendah dengan nilai 1,89 tahun terdapat di Provinsi Bali.

Tabel 4.1 Statistik Deskriptif Variabel Respon dan Variabel Prediktor

| Variabel | Minimum | Maksimum | Range  | Mean   | Standar Deviasi |
|----------|---------|----------|--------|--------|-----------------|
| $y$      | 1,89    | 8,92     | 7,03   | 4,99   | 1,932           |
| $x_1$    | 60,65   | 77,24    | 16,59  | 68,466 | 3,892           |
| $x_2$    | 1,24    | 3,1      | 1,86   | 1,9341 | 0,3842          |
| $x_3$    | 22,53   | 60,67    | 38,14  | 37,2   | 9,13            |
| $x_4$    | 9       | 15478    | 15469  | 719    | 2637            |
| $x_5$    | 0,26    | 105,34   | 105,08 | 18,03  | 23,72           |
| $x_6$    | -0,38   | 9,98     | 10,36  | 5,355  | 1,75            |

Selanjutnya gambaran karakteristik untuk masing-masing variabel prediktor, yaitu tingkat partisipasi angkatan kerja ( $x_1$ ), upah minimum provinsi ( $x_2$ ), persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK ( $x_3$ ), kepadatan penduduk ( $x_4$ ), investasi ( $x_5$ ) dan laju pertumbuhan ekonomi ( $x_6$ ) adalah sebagai berikut :

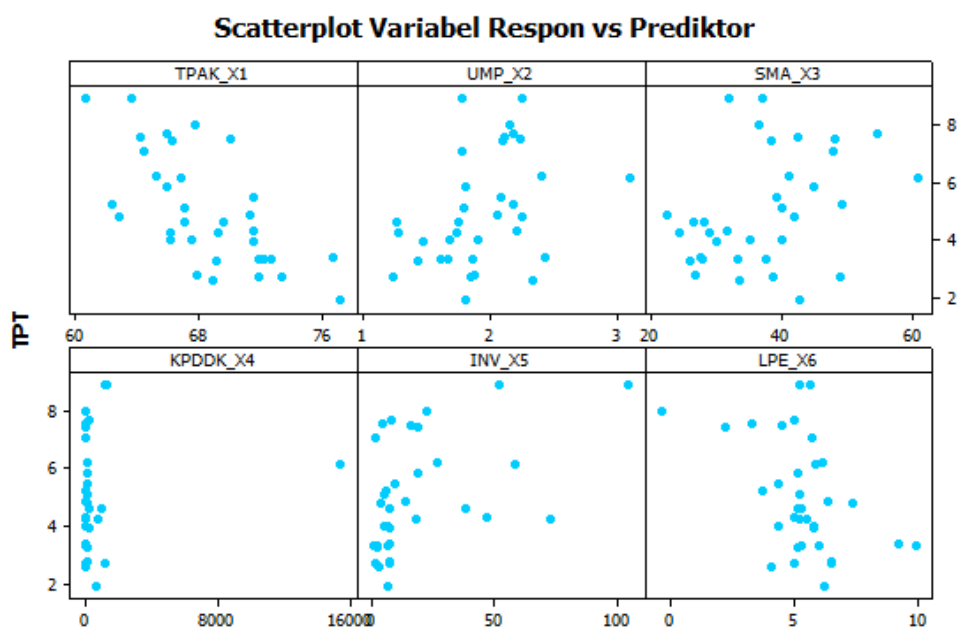
- a. Tingkat partisipasi angkatan kerja di Indonesia pada tahun 2016 yang tertinggi ada di Provinsi Bali dan yang paling rendah ada di Provinsi Jawa Barat, dengan range tingkat partisipasi angkatan kerja adalah sebesar 16,59 persen. Rata-rata tingkat partisipasi angkatan kerja di Indonesia pada tahun 2016 adalah 68,46 persen dengan standar deviasi sebesar 3,89.
- b. Upah minimum provinsi terendah di Indonesia pada tahun 2016 adalah Provinsi DIY yaitu 1,24 juta rupiah dan yang tertinggi adalah Provinsi DKI Jakarta yaitu 3,1 juta rupiah. Sehingga range upah minimum sebesar 1,86 juta rupiah. Rata-rata upah minimum di Indonesia adalah 1,93 juta rupiah dengan standar deviasi sebesar 0,38.
- c. Persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK di Indonesia secara rata-rata pada tahun 2016 adalah 37,2 persen dengan standar deviasi sebesar 9,13. Adapun range persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK di Indonesia adalah 3,18 persen dengan persentase tertinggi adalah Provinsi DKI yaitu 60,67 persen dan persentase yang terendah adalah Provinsi Kalimantan Tengah yaitu 22,53 persen.
- d. Kepadatan penduduk terendah di Indonesia ada di Provinsi Papua Barat dan Provinsi Kalimantan Utara yaitu sebesar 9 jiwa per km<sup>2</sup>, sedangkan kepadatan tertinggi berada di Provinsi DKI yaitu sebesar 15.478 per km<sup>2</sup>. Range kepadatan penduduk 15.649. Rata-rata kepadatan penduduk di Indonesia pada tahun 2016 sebanyak 719 jiwa per km<sup>2</sup>.
- e. Investasi secara rata-rata sebesar 18,03 triliun rupiah. Adapun range besaran Investasi di Indonesia adalah 105,08 triliun rupiah dengan standar deviasi sebesar 23,72. Investasi terendah ada di Provinsi Sulawesi Barat, sedangkan Investasi tertinggi berada di Provinsi Jawa Barat. Laju pertumbuhan ekonomi di Indonesia secara rata rata berada pada angka 5,3 persen, dimana wilayah dengan laju pertumbuhan ekonomi tertinggi di Indonesia pada tahun 2016 berada di Provinsi Sulawesi Tengah yaitu sebesar 9,98 persen dan terendah di Provinsi Kalimantan Timur yang mengalami kontraksi pertumbuhan ekonomi. Dengan standar deviasi antar provinsi sebesar 1.75.

## 4.2.2 Memodelkan TPT Menggunakan Regresi Semiparametrik Campuran Spline *Truncated* dan Kernel

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

### 4.2.2.1 Menentukan Variabel Komponen Parametrik dan Nonparametrik

Tahap awal yang dilakukan untuk melakukan proses pemodelan regresi semiparametrik campuran spline *truncated* dan kernel 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. Dari informasi bentuk pola hubungan tersebut dapat digunakan untuk menentukan variabel prediktor yang akan digunakan sebagai komponen parametrik dan variabel sebagai komponen nonparametrik. Pada Gambar 4.2 menunjukkan bentuk pola hubungan antara variabel TPT dengan variabel-variabel prediktor yaitu tingkat partisipasi angkatan kerja ( $x_1$ ), upah minimum provinsi ( $x_2$ ), persentase penduduk usia 15 tahun keatas



Gambar 4.2 Scatter Plot antara TPT dengan Variabel Prediktor

yang lulus SMA/SMK ( $x_3$ ), kepadatan penduduk ( $x_4$ ), investasi ( $x_5$ ) dan laju pertumbuhan ekonomi ( $x_6$ )

Pola hubungan antara variabel TPT dengan variabel tingkat partisipasi angkatan kerja terlihat membentuk pola linier. Dari hasil *scatter plot* tersebut maka variabel tingkat partisipasi angkatan kerja didekati sebagai komponen parametrik. Sedangkan pola hubungan antara variabel TPT dengan upah minimum provinsi, persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK, kepadatan penduduk, investasi dan laju pertumbuhan ekonomi cenderung tidak membentuk pola tertentu. Pola hubungan yang terlihat cenderung mengalami perubahan perilaku pada sub interval tertentu atau tidak memiliki pola sehingga dapat didekati sebagai komponen nonparametrik.

Selain pendekatan dengan melihat bentuk pola hubungan data melalui scatter plot, untuk menentukan variabel mana yang didekati dengan spline *truncated* atau kernel untuk variabel-variabel nonparametrik, juga dilakukan dengan memeriksa nilai GCV dari masing-masing variabel prediktor. Berdasarkan hasil perhitungan diperoleh nilai GCV untuk semua kemungkinan model campuran spline *truncated* dan kernel seperti terlihat pada Tabel 4.2, Tabel 4.3, Tabel 4.4 dan Tabel 4.5.

Tabel 4.2 GCV dari Model dengan Komponen 1 Parametrik 4 Spline dan 1 Kernel

| No | Variabel   |                      |        | GCV     |         |         | Komb. Knot |
|----|------------|----------------------|--------|---------|---------|---------|------------|
|    | Parametrik | Spline               | Kernel | 1 Knot  | 2 Knot  | 3 Knot  |            |
| 1  | $x_1$      | $x_2, x_3, x_4, x_5$ | $x_6$  | 0,05073 | 0,04922 | 0,04975 | 0,04922    |
| 2  | $x_1$      | $x_2, x_3, x_4, x_6$ | $x_5$  | 0,11620 | 0,06180 | 0,06464 | 0,05867    |
| 3  | $x_1$      | $x_2, x_3, x_5, x_6$ | $x_4$  | 1,93900 | 1,95100 | 2,00200 | 1,93900    |
| 4  | $x_1$      | $x_2, x_4, x_5, x_6$ | $x_3$  | 0,08158 | 0,08162 | 0,08260 | 0,07821    |
| 5  | $x_1$      | $x_3, x_4, x_5, x_6$ | $x_2$  | 0,05996 | 0,05311 | 0,05828 | 0,04997    |

Berdasarkan Tabel 4.2, dicobakan untuk komponen spline terdiri dari 4 variabel dan komponen kernel terdiri dari 1 variabel. Dari semua kemungkinan diperoleh GCV minimum 0.04922 dengan menggunakan 2 titik knot jika komponen splinenya adalah  $x_2, x_3, x_4, x_5$  dan komponen kernelnya adalah  $x_6$ .

Tabel 4.3 GCV dari Model dengan Komponen 1 Parametrik 3 Spline dan 2 Kernel

| No | Variabel    |                 |            | GCV     |         |         | Komb. Knot     |
|----|-------------|-----------------|------------|---------|---------|---------|----------------|
|    | Para metrik | Spline          | Kernel     | 1 Knot  | 2 Knot  | 3 Knot  |                |
| 1  | $x_1$       | $x_2, x_3, x_4$ | $x_5, x_6$ | 0,05455 | 0,05338 | 0,05315 | 0,05011        |
| 2  | $x_1$       | $x_2, x_3, x_5$ | $x_4, x_6$ | 0,07155 | 0,07350 | 0,07597 | 0,06789        |
| 3  | $x_1$       | $x_2, x_3, x_6$ | $x_4, x_5$ | 0,07728 | 0,07269 | 0,08782 | 0,06423        |
| 4  | $x_1$       | $x_2, x_4, x_5$ | $x_3, x_6$ | 0,04763 | 0,05235 | 0,05777 | 0,04763        |
| 5  | $x_1$       | $x_2, x_4, x_6$ | $x_3, x_5$ | 0,06182 | 0,06513 | 0,06606 | 0,06182        |
| 6  | $x_1$       | $x_2, x_5, x_6$ | $x_3, x_4$ | 0,07535 | 0,08523 | 0,09092 | 0,07694        |
| 7  | $x_1$       | $x_3, x_4, x_5$ | $x_2, x_6$ | 0,05510 | 0,04631 | 0,05215 | 0,04636        |
| 8  | $x_1$       | $x_3, x_4, x_6$ | $x_2, x_5$ | 0,05083 | 0,04415 | 0,04544 | <b>0,04054</b> |
| 9  | $x_1$       | $x_3, x_5, x_6$ | $x_2, x_4$ | 0,07871 | 0,08301 | 0,07369 | 0,06455        |
| 10 | $x_1$       | $x_4, x_5, x_6$ | $x_2, x_3$ | 0,05225 | 0,05205 | 0,05391 | 0,04398        |

Berdasarkan Tabel 4.3, dicobakan untuk komponen spline terdiri dari 3 variabel dan komponen kernel terdiri dari 2 variabel. Dari semua kemungkinan diperoleh GCV minimum 0.04054 dengan menggunakan kombinasi titik knot jika komponen splinenya adalah  $x_3, x_4, x_6$  dan komponen kernelnya adalah  $x_2, x_5$ .

Tabel 4.4 GCV dari Model dengan Komponen 1 Parametrik 2 Spline dan 3 Kernel

| No | Variabel    |            |                 | GCV     |         |                | Komb. Knot |
|----|-------------|------------|-----------------|---------|---------|----------------|------------|
|    | Para metrik | Spline     | Kernel          | 1 Knot  | 2 Knot  | 3 Knot         |            |
| 1  | $x_1$       | $x_2, x_3$ | $x_4, x_5, x_6$ | 0,05884 | 0,05796 | 0,06143        | 0,05508    |
| 2  | $x_1$       | $x_2, x_4$ | $x_3, x_5, x_6$ | 0,05073 | 0,04922 | 0,04975        | 0,04922    |
| 3  | $x_1$       | $x_2, x_5$ | $x_3, x_4, x_6$ | 0,05918 | 0,06433 | 0,06665        | 0,06005    |
| 4  | $x_1$       | $x_2, x_6$ | $x_3, x_4, x_5$ | 0,06213 | 0,06648 | 0,06630        | 0,06346    |
| 5  | $x_1$       | $x_3, x_4$ | $x_2, x_5, x_6$ | 0,05173 | 0,04733 | 0,05218        | 0,04733    |
| 6  | $x_1$       | $x_3, x_5$ | $x_2, x_4, x_6$ | 0,06061 | 0,06187 | 0,06053        | 0,05973    |
| 7  | $x_1$       | $x_3, x_6$ | $x_2, x_4, x_5$ | 0,06150 | 0,06365 | 0,06755        | 0,06150    |
| 8  | $x_1$       | $x_4, x_5$ | $x_2, x_3, x_6$ | 0,05661 | 0,04567 | 0,05669        | 0,04540    |
| 9  | $x_1$       | $x_4, x_6$ | $x_2, x_3, x_5$ | 0,04703 | 0,04427 | <b>0,04299</b> | 0,04299    |
| 10 | $x_1$       | $x_4, x_6$ | $x_2, x_3, x_4$ | 0,06525 | 0,06641 | 0,07227        | 0,06598    |



Berdasarkan Tabel 4.4, dicobakan untuk komponen spline terdiri dari 2 variabel dan komponen kernel terdiri dari 3 variabel. Dari semua kemungkinan diperoleh GCV minimum 0.04299 dengan menggunakan 3 titik knot jika komponen splinenya adalah  $x_4, x_6$  dan komponen kernelnya adalah  $x_2, x_3, x_5$ .

Tabel 4.5 GCV dari Model dengan Komponen 1 Parametrik 1 Spline dan 4 Kernel

| No | Para metrik | Variabel |                      | GCV     |         |         |
|----|-------------|----------|----------------------|---------|---------|---------|
|    |             | Spline   | Kernel               | 1 Knot  | 2 Knot  | 3 Knot  |
| 1  | $x_1$       | $x_2$    | $x_3, x_4, x_5, x_6$ | 0,05672 | 0,05728 | 0,05711 |
| 2  | $x_1$       | $x_3$    | $x_2, x_4, x_5, x_6$ | 0,05505 | 0,05505 | 0,05617 |
| 3  | $x_1$       | $x_4$    | $x_2, x_3, x_5, x_6$ | 0,04988 | 0,05008 | 0,05008 |
| 4  | $x_1$       | $x_5$    | $x_2, x_3, x_4, x_6$ | 0,05489 | 0,05576 | 0,05647 |
| 5  | $x_1$       | $x_6$    | $x_2, x_3, x_4, x_5$ | 0,05561 | 0,05843 | 0,05931 |

Berdasarkan Tabel 4.5, dicobakan untuk komponen spline terdiri dari 1 variabel dan komponen kernel terdiri dari 4 variabel. Dari semua kemungkinan diperoleh GCV minimum 0,04988 dengan menggunakan 1 titik knot jika komponen splinenya adalah  $x_4$  dan komponen kernelnya adalah  $x_2, x_3, x_5, x_6$ .

Hasil semua kombinasi variabel untuk komponen spline dan komponen kernel yang telah dicobakan, diperoleh GCV minimum jika komponen splinenya adalah  $x_3, x_4, x_6$  dan komponen kernelnya adalah  $x_2, x_5$  dengan nilai GCV sebesar 0,04054 dengan menggunakan kombinasi knot yang merupakan nilai GCV terkecil diantara semua kemungkinan yang dicobakan.

Dengan demikian data TPT didekati dengan regresi semiparametrik campuran spline truncated dan kernel, dimana terdapat 1 variabel prediktor yang didekati dengan parametrik yaitu variabel tingkat partisipasi angkatan kerja, sementara kelima variabel lainnya didekati dengan nonparametrik. Pada komponen nonparametrik, variabel persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK, kepadatan penduduk, dan laju pertumbuhan ekonomi didekati dengan fungsi spline *truncated*, sementara itu, dan variabel upah minimum provinsi dan variabel investasi didekati dengan fungsi kernel. Selanjutnya variabel yang didekati

dengan parametrik linear disimbolkan dengan ( $x$ ), variabel yang didekati dengan fungsi spline *truncated* disimbolkan dengan ( $t$ ) dan variabel yang didekati dengan fungsi kernel disimbolkan dengan ( $z$ ). Ringkasan hasil penentuan komponen parametrik dan komponen nonparametrik dituliskan pada Tabel 4.6.

Tabel 4.6 Ringkasan Penentuan Komponen Parametrik dan Nonparametrik

| Notasi | Nama Variabel   | Komponen   |
|--------|---|------------|
| $x_1$  | Tingkat Partisipasi Angkatan Kerja                          | Parametrik |
| $t_1$  | Persentase Penduduk Usia 15 tahun keatas yang Lulus SMA/SMK | Spline     |
| $t_2$  | Kepadatan Penduduk  | Spline     |
| $t_3$  | Laju Pertumbuhan Ekonomi                                    | Spline     |
| $z_1$  | Upah Minimum Provinsi                                       | Kernel     |
| $z_2$  | Investasi   | Kernel     |

#### 4.2.2.2 Memodelkan Tingkat Pengangguran Terbuka dengan Satu Titik Knot

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

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{13} t_{3i} + \theta_{23} (t_{3i} - K_{13})_+^1 + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}} \right) y_i \varepsilon_i \quad i = 1, 2, \dots, n$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan satu knot disajikan dalam Tabel 4.7.

Tabel 4.7 Perbandingan Nilai GCV dengan menggunakan Satu Titik Knot

| Spline         |                |                | Kernel     |            | GCV      |
|----------------|----------------|----------------|------------|------------|----------|
| Knot           |                |                | Bandwidth  |            |          |
| $t_1 = K_{11}$ | $t_2 = K_{12}$ | $t_3 = K_{13}$ | $\alpha_1$ | $\alpha_2$ |          |
| 51,58905       | 11794,9        | 7,513333       | 0,033518   | 2,665736   | 0,050827 |
| 53,40524       | 12531,52       | 8,006667       | 0,033654   | 2,661295   | 0,051331 |
| 49,77286       | 11058,29       | 7,02           | 0,382589   | 2,973183   | 0,052383 |
| 55,22143       | 13268,14       | 8,5            | 0,038312   | 2,81849    | 0,053269 |

Berdasarkan Tabel 4.7 nilai GCV minimum yang dihasilkan adalah sebesar 0,05082. Lokasi titik knot pada variabel ( $t_1$ ) yaitu 51,5890 ( $K_{11}$ ), ( $t_2$ ) yaitu 11,794 ( $K_{12}$ ), dan ( $t_3$ ) yaitu 7,513 ( $K_{13}$ ), sedangkan bandwidth yang diberikan sebesar  $\alpha_1 = 0,3315$  dan  $\alpha_2 = 2,6657$ .

#### 4.2.2.3 Memodelkan Tingkat Pengangguran Terbuka 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:

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_i + \theta_{21} (t_i - K_{11})_+^1 + \theta_{31} (t_i - K_{21})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{32} (t_{2i} - K_{22})_+^1 + \theta_{13} t_{3i} + \theta_{23} (t_{3i} - K_{13})_+^1 + \theta_{33} (t_{3i} - K_{23})_+^1 + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}} \right) y_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan satu knot disajikan dalam Tabel 4.8

Tabel 4.8 Perbandingan Nilai GCV dengan menggunakan Dua Titik Knot

| Spline         |                |                | Kernel     |            | GCV      |
|----------------|----------------|----------------|------------|------------|----------|
| Knot           |                |                | Bandwidth  |            |          |
| $t_1 = K_{11}$ | $t_2 = K_{12}$ | $t_3 = K_{13}$ | $\alpha_1$ | $\alpha_2$ |          |
| $t_1 = K_{21}$ | $t_1 = K_{22}$ | $t_1 = K_{23}$ |            |            |          |
| 48,93462       | 10718,31       | 6,792308       | 0,03672    | 0,112614   | 0,044151 |
| 51,86846       | 11908,23       | 7,589231       |            |            |          |
| 51,86846       | 11908,23       | 7,589231       | 0,012485   | 3,126704   | 0,052597 |
| 57,73615       | 14288,08       | 9,183077       |            |            |          |
| 51,86846       | 11908,23       | 7,589231       | 0,012485   | 3,126703   | 0,052597 |
| 54,80231       | 13098,15       | 8,386154       |            |            |          |
| 48,93462       | 10718,31       | 6,792308       | 0,420904   | 3,070184   | 0,057437 |
| 57,73615       | 14288,08       | 9,183077       |            |            |          |

Berdasarkan Tabel 4.8 nilai GCV minimum yang dihasilkan adalah sebesar 0,044151. Lokasi titik knot pada variabel ( $t_1$ ) yaitu 48,9346 ( $K_{11}$ ), 51,8684 ( $K_{21}$ ),

untuk  $(t_2)$  yaitu 10718,31 ( $K_{12}$ ) 11908,23 ( $K_{22}$ ) untuk untuk  $(t_3)$  yaitu 6,7923( $K_{13}$ ) 7,5892 ( $K_{23}$ ) sedangkan bandwidth yang diberikan sebesar  $\alpha_1 = 0,03672$  ,  $\alpha_2 = 0,1126$ .

#### 4.2.2.4 Memodelkan Tingkat Pengangguran Terbuka 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.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{31} (t_{1i} - K_{21})_+^1 + \theta_{41} (t_{1i} - K_{31})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{32} (t_{2i} - K_{22})_+^1 + \theta_{42} (t_{2i} - K_{32})_+^1 + \theta_{13} t_{3i} + \theta_{23} (t_{3i} - K_{13})_+^1 + \theta_{33} (t_{3i} - K_{23})_+^1 + \theta_{43} (t_{3i} - K_{33})_+^1 + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}} \right) y_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan tiga knot disajikan dalam Tabel 4.9.

Tabel 4.9 Perbandingan Nilai GCV dengan menggunakan Tiga Titik Knot

| SplineKnot     |                |                | Kernel     |            | GCV      |
|----------------|----------------|----------------|------------|------------|----------|
| Knot           |                |                | Bandwidth  |            |          |
| $t_1 = K_{11}$ | $t_2 = K_{12}$ | $t_3 = K_{13}$ | $\alpha_1$ | $\alpha_2$ |          |
| $t_1 = K_{21}$ | $t_2 = K_{22}$ | $t_3 = K_{23}$ |            |            |          |
| $t_1 = K_{31}$ | $t_2 = K_{22}$ | $t_3 = K_{33}$ |            |            |          |
| 43,33364       | 8446,636       | 5,270909       |            |            |          |
| 46,80091       | 9852,909       | 6,212727       | 0,037743   | 0,279557   | 0.045439 |
| 50,26818       | 11259,18       | 7,154545       |            |            |          |
| 43,33364       | 8446,636       | 5,270909       |            |            |          |
| 46,80091       | 9852,909       | 6,212727       | 0,037557   | 0,279859   | 0.052084 |
| 53,73545       | 12665,45       | 8,096364       |            |            |          |
| 50,26818       | 11259,18       | 7,154545       |            |            |          |
| 53,73545       | 12665,45       | 8,096364       | 0,012458   | 3,173306   | 0.055339 |
| 57,20273       | 14071,73       | 9,038182       |            |            |          |
| 46,80091       | 9852,909       | 6,212727       |            |            |          |
| 50,26818       | 11259,18       | 7,154545       | 0,039134   | 0,482074   | 0.056661 |
| 57,20273       | 14071,73       | 9,038182       |            |            |          |

Berdasarkan Tabel 4.9 nilai GCV minimum yang dihasilkan adalah sebesar 0,0454 Lokasi titik knot pada variabel ( $t_1$ ) yaitu 43,33364 ( $K_{11}$ ) 46,80091 ( $K_{21}$ ) 50,26818( $K_{31}$ ) untuk ( $t_2$ ) yaitu 8446,636 ( $K_{12}$ ) 9852,909 ( $K_{22}$ ) 11259,18 ( $K_{32}$ ) dan untuk ( $t_3$ ) yaitu 5,270909 ( $K_{13}$ ) 6,212727 ( $K_{23}$ ) 7,154545 ( $K_{33}$ ) sedangkan bandwidth yang diberikan sebesar  $\alpha_1 = 0,037743$  dan  $\alpha_2 = 0,279557$ .

#### 4.2.2.5 Memodelkan Tingkat Pengangguran Terbuka 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 empat komponen nonparametrik adalah sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \theta_{11} t_{1i} + \theta_{21} (t_{1i} - K_{11})_+^1 + \theta_{31} (t_{1i} - K_{21})_+^1 + \theta_{12} t_{2i} + \theta_{22} (t_{2i} - K_{12})_+^1 + \theta_{32} (t_{2i} - K_{22})_+^1 + \theta_{13} t_{3i} + \theta_{23} (t_{3i} - K_{13})_+^1 + \theta_{33} (t_{3i} - K_{23})_+^1 + \theta_{43} (t_{3i} - K_{33})_+^1 + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{\alpha_1} \right)^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}}{\sum_{i=1}^n \frac{1}{\alpha_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{\alpha_2} \right)^2}} \right) y_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Nilai GCV yang dihasilkan dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel dengan kombinasi knot disajikan dalam Tabel 4.10.

Berdasarkan Tabel 4.10 nilai GCV minimum yang dihasilkan adalah sebesar 0,0454 dengan kombinasi titik knot, 2,2,3 dan lokasi titik knot pada variabel ( $t_1$ ) yaitu 48.935 ( $K_{11}$ ) 51.868 ( $K_{21}$ ) untuk ( $t_2$ ) yaitu 10718.31 ( $K_{12}$ ) 11908.23 ( $K_{22}$ ) dan untuk ( $t_3$ ) yaitu 5.271 ( $K_{13}$ ) 6.213 ( $K_{23}$ ) 7.155 ( $K_{33}$ ) sedangkan bandwidth yang diberikan sebesar  $\alpha_1 = 0.037642$  dan  $\alpha_2 = 0.259689$ .

Tabel 4.10 Perbandingan Nilai GCV dengan menggunakan Kombinasi Titik Knot

| SplineKnot     |                |                | Kernel     |            | GCV      | Kombinasi |
|----------------|----------------|----------------|------------|------------|----------|-----------|
| Knot           |                |                | Bandwidth  |            |          |           |
| $t_1 = K_{11}$ | $t_2 = K_{12}$ | $t_3 = K_{13}$ | $\alpha_1$ | $\alpha_2$ |          |           |
| $t_1 = K_{21}$ | $t_2 = K_{22}$ | $t_3 = K_{23}$ |            |            |          |           |
| $t_1 = K_{31}$ | $t_2 = K_{22}$ | $t_3 = K_{33}$ |            |            |          |           |
| 48.935         | 10718.31       | 5.271          |            |            |          |           |
| 51.868         | 11908.23       | 6.213          | 0.037642   | 0.259689   | 0.040543 | 2,2,3     |
|                |                | 7.155          |            |            |          |           |
| 48.935         | 11794.91       | 5.271          |            |            |          |           |
| 51.868         |                | 6.213          | 0.037642   | 0.259689   | 0.040543 | 2,1,3     |
|                |                | 7.155          |            |            |          |           |
| 48.935         | 11794.91       | 6.792          |            |            |          |           |
| 51.868         |                | 7.589          | 0.03672    | 0.112614   | 0.044151 | 2,1,2     |
|                |                |                |            |            |          |           |
| 43.334         | 11794.91       | 5.271          |            |            |          |           |
| 46.801         |                | 6.213          | 0.037743   | 0.279557   | 0.045439 | 3,1,3     |
| 50.268         |                | 7.155          |            |            |          |           |

#### 4.2.2.6 Estimasi Interval Parameter Model Semiparametrik Campuran Spline *Truncated* dan Kernel

Berdasarkan model semiparametrik campuran spline *truncated* dan kernel terbaik untuk TPT di Indonesia, selanjutnya akan di hitung estimasi interval untuk parameter model semiparametrik campuran spline *truncated* dan kernel. Dari hasil yang telah didapatkan pada subbab 4.14 akan didapatkan estimasi interval untuk parameter model dengan tingkat kepercayaan sebesar 95% seperti persamaan (4.73)

Estimasi interval untuk parameter dapat digunakan untuk mengetahui signifikansi parameter terhadap model. Pengambilan kesimpulan tersebut dilakukan dengan melihat apakah interval parameter memuat nilai nol. Jika interval memuat nilai nol, maka parameter tidak berpengaruh secara signifikan terhadap model. Berikut ini adalah hasil estimasi interval untuk parameter model regresi semiparametrik campuran spline *truncated* dan kernel.

Berdasarkan Tabel 4.11 memperlihatkan bahwa dari 12 parameter pada model regresi semiparametrik campuran spline *truncated* dan kernel, terdapat 11 parameter yang signifikan. Secara keseluruhan semua variabel pada komponen parametrik dan komponen spline *truncated* yang diduga berpengaruh signifikan

terhadap TPT yaitu : tingkat partisipasi angkatan kerja, persentase penduduk usia 15 tahun keatas yang lulus SMA/SMK, kepadatan penduduk, dan laju pertumbuhan ekonomi.

Tabel 4.11 Estimasi Interval untuk Parameter Komponen Parametrik dan Spline dalam Model Semiparametrik Campuran Spline *Truncated* dan Kernel

| Variabel                                  | Parameter     | Estimasi  | Batas Bawah | Batas Atas | Keputusan  |
|---|---------------|-----------|-------------|------------|------------|
| -   | $\beta_0$     | 11.40312  | 12.44424    | 10.362     | Signifikan |
| $(x_1)$<br>TPAK                           | $\beta_1$     | -0.17128  | -0.1561     | -0.18646   | Signifikan |
| $(t_1)$                                   | $\theta_{11}$ | 0.018691  | 0.024029    | 0.013352   | Signifikan |
| Persentase penduduk 15 +<br>Lulus SMA/SMK | $\theta_{21}$ | -7.80463  | -6.33439    | -9.27488   | Signifikan |
|   | $\theta_{31}$ | 16.3012   | 19.27254    | 13.32987   | Signifikan |
| $(t_2)$<br>Kepadatan Penduduk             | $\theta_{12}$ | -8.27E-06 | 0.000161    | -0.00018   | Tidak      |
|   | $\theta_{22}$ | -0.00708  | -0.00582    | -0.00835   | Signifikan |
|   | $\theta_{32}$ | -0.00531  | -0.00436    | -0.00626   | Signifikan |
| $(t_3)$<br>Laju pertumbuhan<br>Ekonomi    | $\theta_{13}$ | -0.06905  | -0.02199    | -0.1161    | Signifikan |
|   | $\theta_{23}$ | 0.422929  | 0.662729    | 0.18313    | Signifikan |
|   | $\theta_{43}$ | -2.8226   | -2.2284     | -3.41679   | Signifikan |
|   | $\theta_{43}$ | 3.474633  | 4.07565     | 2.873615   | Signifikan |

Estimasi interval pada parameter memuat informasi yang lebih dibandingkan dengan estimasi titik pada pengujian hipotesis, karena memiliki batas atas dan batas bawah pada setiap interval yang dihasilkan. Sehingga nilai-nilai yang berada dalam interval menghasilkan kesimpulan yang sama. Selain itu, untuk mendapatkan kesimpulan dari estimasi interval lebih sederhana dibandingkan pengujian hipotesis, baik dari tahapan maupun penghitungan matematisnya. Oleh karena itu estimasi interval dengan pendekatan *pivotal quantity* digunakan dalam penelitian ini.

#### 4.2.2.7 Model Semiparametrik Campuran Spline Truncated dan Kernel dengan Titik Knot dan Bandwidth Optimum

Setelah didapatkan model regresi semiparametrik spline *truncated* dengan satu titik knot, dua titik knot, tiga titik knot dan kombinasi knot, maka langkah

selanjutnya adalah memilih model terbaik dengan membandingkan nilai GCV pada masing-masing model. Nilai GCV minimum pada masing-masing model disajikan pada Tabel 4.12.

Tabel 4.12 Perbandingan Nilai GCV Minimum Tiap Model

| Banyaknya Knot | GCV     |
|----------------|---------|
| 1 Knot         | 0,05083 |
| 2 Knot         | 0,04415 |
| 3 Knot         | 0,04544 |
| Kombinasi Knot | 0,04054 |

Berdasarkan Tabel 4.12 terlihat bahwa GCV minimum terdapat pada kombinasi knot dengan nilai GCV sebesar 0,04054. Model regresi semiparametrik campuran spline *truncated* dan kernel terbaik yaitu dengan kombinasi titik knot (2,2,3). Dari hasil estimasi parameter, titik knot dan bandwidth optimum yang diperoleh maka model yang terbentuk adalah sebagai berikut.

$$\hat{y}_i = 11,4031 - 0,1711x_i - 0,01869t_{1i} - 7,8046(t_{1i} - 48,935)_+^1 + 16,3012(t_{1i} - 51,868)_+^1 +$$

$$- 0,000008t_{2i} - 0,0070(t_{2i} - 10718,31)_+^1 - 0,0053(t_{2i} - 11908,23)_+^1 - 0,069t_{3i} +$$

$$0,4229(t_{3i} - 5,271)_+^1 - 2,822(t_{3i} - 6,213)_+^1 + 3,4746(t_{3i} - 7,155)_+^1 +$$

$$\sum_{i=1}^n \left( \frac{\frac{1}{0,0376} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{0,0376} \right)^2}}{\sum_{i=1}^n \frac{1}{0,0376} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{0,0376} \right)^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{0,2596} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{0,2596} \right)^2}}{\sum_{i=1}^n \frac{1}{0,2596} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{0,2596} \right)^2}} \right) y_i$$

Dari model yang dihasilkan akan disajikan beberapa contoh interpretasi model terhadap variabel-variabel yang berpengaruh secara signifikan. Berikut penjelasan tentang interpretasi model dan gambaran perubahan pola perilaku pada beberapa variabel.

1. Interpretasi terhadap TPAK ( $x_1$ ) dengan asumsi variabel lain konstan adalah sebagai berikut:

$$\hat{y} = 11,4031 - 0,1711x_1$$

Jika terjadi kenaikan TPAK sebanyak satu persen maka nilai TPT akan turun sebesar 0,17 persen. Hasil pemodelan tersebut sesuai dengan hipotesis awal, yaitu semakin bertambah TPAK maka TPT akan semakin menurun. Dengan

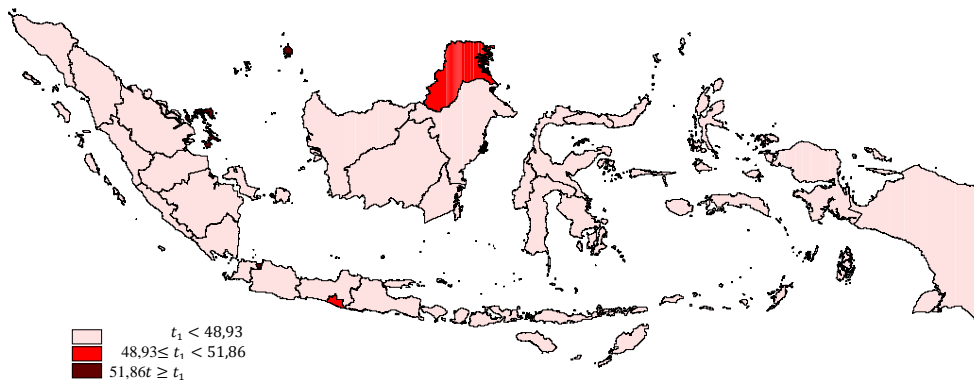


indikasi arah hubungan antara TPAK dengan TPT adalah negatif. Dengan meningkatnya TPAK dan tersedianya lapangan kerja dapat menurunkan TPT di Indonesia.

2. Interpretasi terhadap variabel pendidikan yaitu persentase penduduk usia 15+ yang lulus SMA/SMK ( $t_1$ ) dengan asumsi variabel lain konstan adalah sebagai berikut.

$$\hat{y} = \begin{cases} -0,0186t_1 & ; & t_1 < 48,93 \\ -7,82329t_1 + 381,918 & ; & 48,935 \leq t_1 < 51,86 \\ 8,47791t_1 - 463,592 & ; & t_1 \geq 51,86 \end{cases}$$

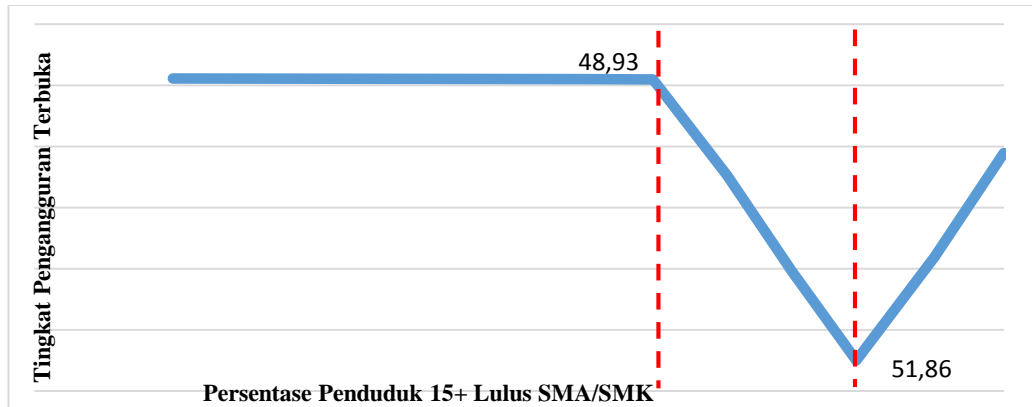
Dengan regresi semiparametrik campuran spline *truncated* dan kernel maka variabel yang mempunyai pola spline *truncated* dapat diinterpretasikan sesuai dengan pola perubahan datanya atau pada setiap interval knotnya. Pada variabel persentase penduduk 15+ yang lulus SMA/SMK memiliki 3 sub interval perubahan perilaku. Untuk nilai kurang dari 48,93 persen, setiap kenaikan satu persen maka TPT akan mengalami penurunan sebesar 0,018 persen. Dari model yang dihasilkan sebagian besar provinsi di Indonesia mengikuti pola seperti ini. Adapun pembagian provinsi sesuai dengan pola perilaku datanya terlihat pada Gambar 4.3.



Gambar 4.3 Peta TPT terhadap Persentase Penduduk 15+ yang lulus SMA/SMK

Berikut disajikan visualisasi perubahan pola perilaku pada variabel persentase penduduk 15+ yang lulus SMA/SMK. Berdasarkan Gambar 4.4 terlihat perubahan perilaku pada interval  $t_1 < 48,93$  menyebabkan penurunan TPT secara lambat, sedangkan pada interval  $48,935 \leq t_1 < 51,86$  masih terjadi penurunan

terhadap TPT tetapi dengan pola perilaku yang terjadi lebih tajam. Untuk interval  $t_1 \geq 51,86$  pola perilaku berubah arah dibandingkan dengan dua interval sebelumnya yaitu terjadi kenaikan pada TPT seiring dengan semakin bertambahnya persentase penduduk 15+ yang lulus SMA/SMK.



Gambar 4.4 Hubungan antara persentase penduduk 15+ yang lulus SMA/SMK dengan TPT

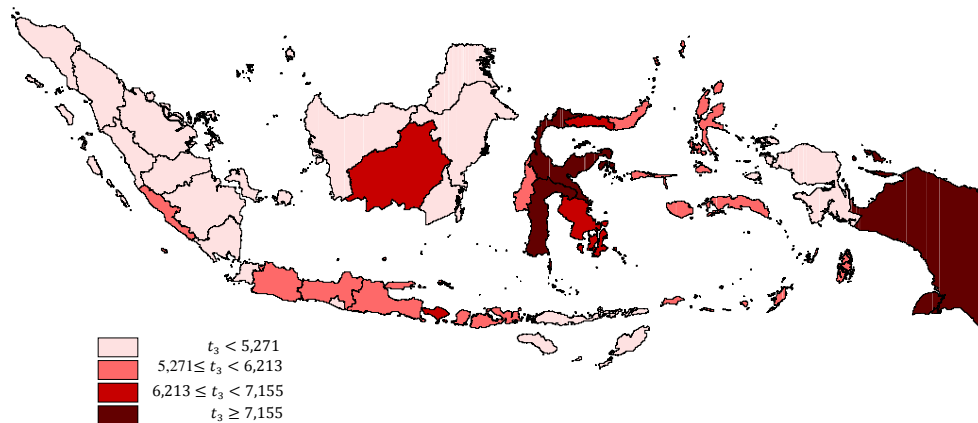
3. Interpretasi terhadap variabel laju pertumbuhan ekonomi, ( $t_3$ ) dengan asumsi variabel lain konstan adalah sebagai berikut.

$$\hat{y} = \begin{cases} -0,069t_3 & ; t_3 < 5,271 \\ 0,354t_3 - 2,229 & ; 5,271 \leq t_3 < 6,213 \\ -2,468t_3 + 15,304 & ; 6,213 \leq t_3 < 7,155 \\ 1,0065t_3 - 9,556 & ; t_3 \geq 7,155 \end{cases}$$

Pada variabel laju pertumbuhan ekonomi terdapat empat sub interval perubahan perilaku. Untuk provinsi dengan laju pertumbuhan ekonomi kurang dari 5,271 persen, terjadi hubungan negatif antara laju pertumbuhan ekonomi dengan TPT. Setiap kenaikan satu persen laju pertumbuhan ekonomi akan menyebabkan penurunan TPT sebesar 0,069 persen. Sedangkan provinsi yang memiliki laju pertumbuhan ekonomi antara 6,213 persen hingga 7,155 juga mempunyai hubungan negatif dengan pola perilaku yang lebih tajam.

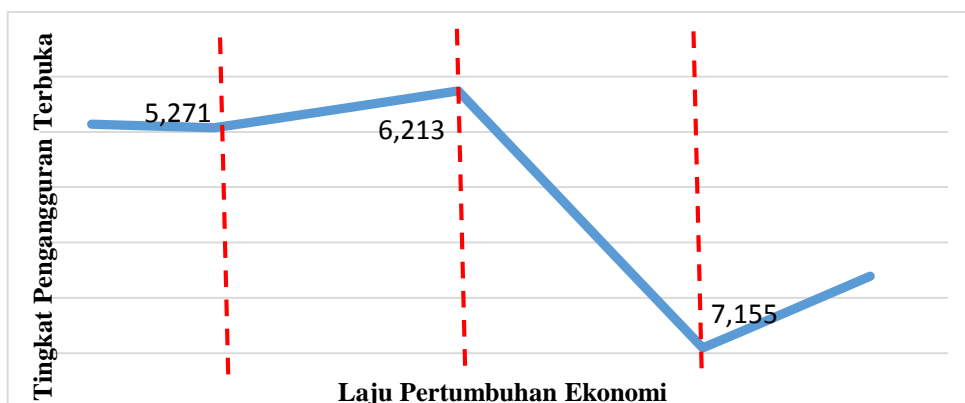
Dari model yang dihasilkan ada sebanyak 27 provinsi atau sebesar 79 persen yang mengikuti pola perilaku seperti ini. Adapun sebanyak 21 persen atau 7

provinsi yang memiliki pola perilaku yang sebaliknya. Adapun pembagian provinsi sesuai dengan pola perilaku datanya terlihat pada Gambar 4.3.



Gambar 4.5 Peta TPT terhadap Laju Pertumbuhan Ekonomi

Berikut disajikan visualisasi perubahan pola perilaku pada variabel laju pertumbuhan ekonomi. Berdasarkan Gambar 4.6 terlihat perubahan perilaku pada interval  $t_3 < 5,271$  menyebabkan penurunan TPT secara lambat, sedangkan pada interval  $6,213 \leq t_3 < 7,155$  masih terjadi penurunan terhadap TPT tetapi dengan pola perilaku yang terjadi lebih tajam. Untuk interval  $5,271 \leq t_3 < 6,213$  dan  $t_3 \geq 7,155$  pola perilaku berubah arah dibandingkan dengan dua interval sebelumnya yaitu terjadi kenaikan pada TPT seiring dengan semakin bertambahnya Laju Pertumbuhan Ekonomi.



Gambar 4.6 Hubungan Antara Laju Pertumbuhan Ekonomi terhadap TPT

### 4.2.3 Pemeriksaan 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*. Pendeteksian asumsi independen menggunakan plot *Autocorrelation Function* (ACF). Dan pengujian asumsi normal menggunakan *Kolmogorov-Smirnov*.

#### 4.2.3.1 Pemeriksaan Asumsi Identik

Pengujian asumsi identik dimaksudkan untuk mendeteksi apakah varians dari residualnya memiliki varians yang homogen atau tidak. Pemeriksaan menggunakan uji *glejser*. Hipotesis yang digunakan adalah:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{paling sedikit terdapat satu } \sigma_i^2 \neq \sigma^2 ; \text{ dimana } i = 1, 2, \dots, n$$

Hasil pengujian *glejser* tersaji pada Tabel 4.13.

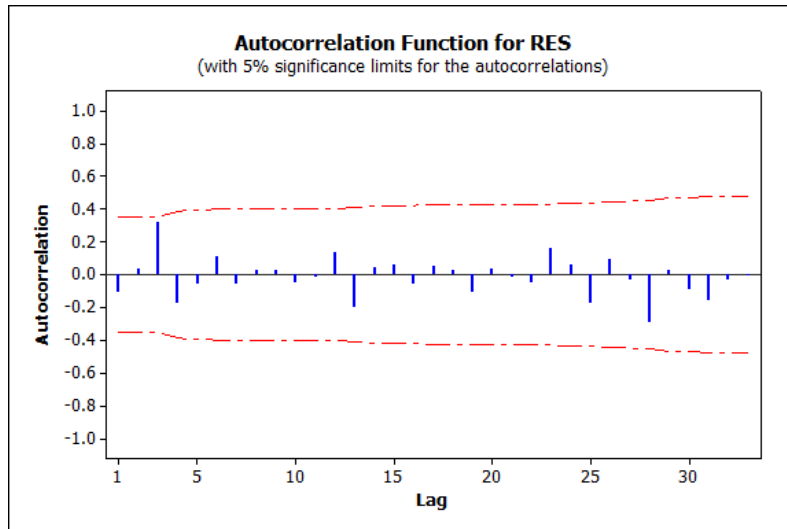
Tabel 4.13 Hasil Uji *Glejser*

| Sumber Variasi | Derajat Kebebasan | Jumlah Kuadrat | Rata-rata Kuadrat | F <sub>hitung</sub> | p-value |
|----------------|-------------------|----------------|-------------------|---------------------|---------|
| Regresi        | 10                | 0,78352        | 0,07835           | 1,75                | 0,130   |
| Error          | 23                | 1,03169        | 0,04486           |                     |         |
| Total          | 33                | 1,81521        |                   |                     |         |

Dari Tabel 4.13 terlihat bahwa nilai *p-value* sebesar 0,130 lebih besar dari nilai  $\alpha$  yang ditetapkan yaitu 0,05 sehingga keputusan yang diambil adalah gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa varians dari residual adalah homogen, sehingga asumsi identik pada model regresi semiparametrik campuran spline *truncated* dan kernel yang terbentuk telah terpenuhi.

#### 4.2.3.2 Pemeriksaan Asumsi Independen

Pemeriksaan asumsi independen dimaksudkan untuk mendeteksi apakah terjadi korelasi antar residual. Pemeriksaan dilakukan dengan menggunakan plot ACF. Hasil dari plot ACF tersaji pada Gambar 4.7.



Gambar 4.7 Plot Autocorrelation Function (ACF) Residual

Terlihat bahwa tidak ada lag yang keluar dari batas toleransi, hal ini menandakan bahwa tidak ada korelasi antar residual. Berdasarkan hasil pemeriksaan plot ACF residual, maka asumsi Independen pada model regresi semiparametrik spline *truncated* yang terbentuk telah terpenuhi.

#### 4.2.3.3 Pemeriksaan Asumsi Distribusi Normal

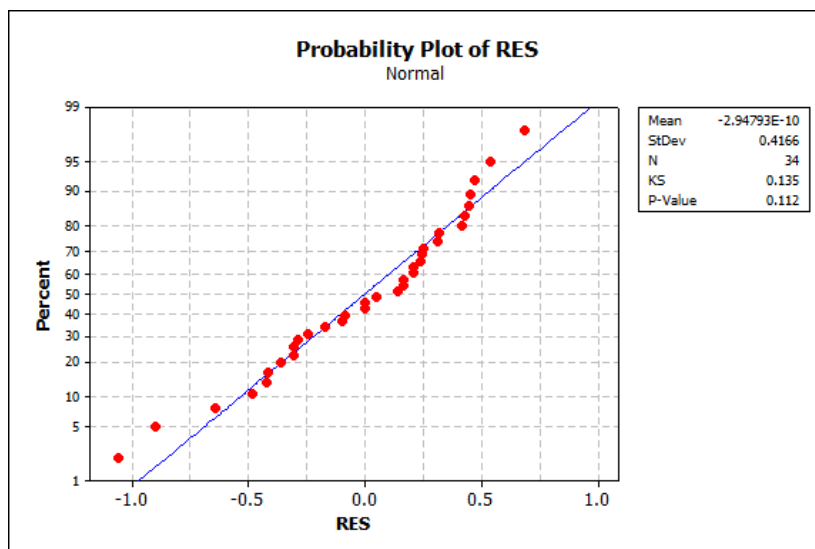
Residual regresi semiparametrik campuran spline *truncated* dan kernel harus memenuhi asumsi distribusi normal. Oleh karenanya dilakukan pemeriksaan asumsi distribusi normal pada residual yang diperoleh. Pemeriksaan dilakukan dengan menggunakan uji *Kolmogorov-Smirnov*. Hipotesis yang digunakan adalah sebagai berikut:

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

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

Hasil uji *Kolmogorov-Smirnov* disajikan dalam bentuk plot seperti terlihat pada Gambar 4.8. Plot residual seperti yang terlihat pada Gambar 4.8 menghasilkan *p-value* sebesar 0,112 yang lebih besar dari nilai  $\alpha$  yang ditentukan yaitu 0,05, sehingga keputusan yang diambil adalah gagal tolak  $H_0$ . Oleh karenanya dapat disimpulkan bahwa residual berdistribusi normal. Dari hasil uji ini

maka regresi semiparametrik campuran spline *truncated* dan kernel yang dihasilkan memenuhi asumsi berdistribusi normal.



Gambar 4.8 Probability Plot Residual

#### 4.2.4 Perbandingan Model

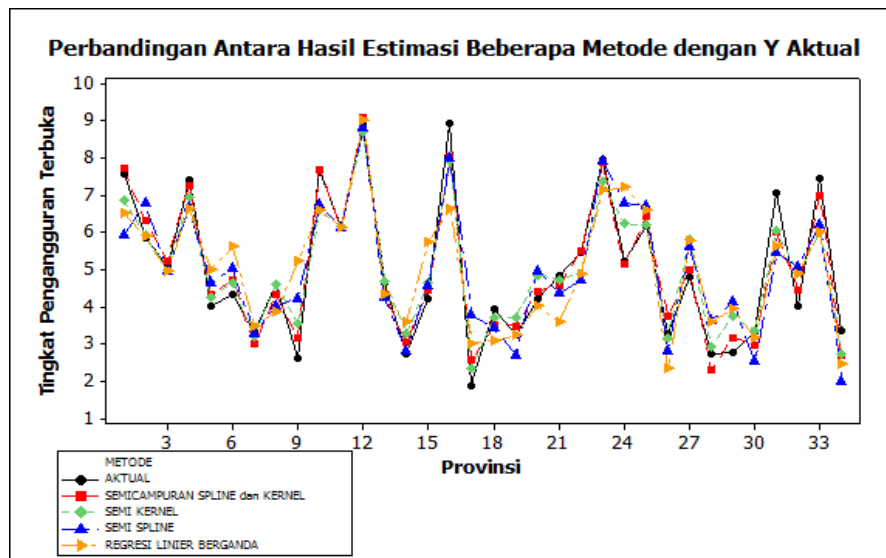
Setelah dilakukan analisis menggunakan metode analisis regresi semiparametrik campuran spline *truncated* dan kernel, selanjutnya akan dilakukan perbandingan dengan beberapa model. Kriteria dalam pemilihan model terbaik menggunakan koefisien determinasi yang disesuaikan ( $R^2$  adjusted). Penggunaan  $R^2$  adjusted karena antara metode memiliki jumlah parameter yang berbeda. Berikut ini adalah nilai  $R^2$  adjusted pada masing-masing model:

Pada Tabel 4.14 terlihat bahwa nilai  $R^2$  adjusted pada model regresi semiparametrik campuran spline *truncated* dan kernel menghasilkan nilai yang lebih besar yaitu 92,53 persen.

Tabel 4.14 Nilai  $R^2$  adjusted

| Model   | $R^2$ adjusted (%) |
|---|--------------------|
| Regresi Linier Berganda                                   | 52,82              |
| Regresi Semiparametrik Spline <i>Truncated</i>            | 62,63              |
| Regresi Semiparametrik Kernel                             | 85,40              |
| Regresi Semiparametrik Spline <i>Truncated</i> dan Kernel | 92,53              |

Dari nilai  $R^2$  adjusted yang diperoleh sebesar 92,53 persen maka model layak untuk digunakan. Oleh karena itu metode yang dipilih adalah metode regresi semiparametrik campuran spline *truncated* dan kernel. Berikut akan ditampilkan perbandingan hasil estimasi metode semiparametrik campuran spline *truncated* dan kernel dengan beberapa metode yang dapat dilihat seperti pada Gambar 4.9 berikut.



Gambar 4.9 Perbandingan Hasil Estimasi beberapa Metode

Dari Gambar 4.9 terlihat metode dengan menggunakan regresi semiparametrik campuran spline *truncated* dan kernel mampu memberikan hasil estimasi yang lebih baik daripada metode yang lain.

Model tersebut telah memenuhi asumsi residual IIDN, sehingga model yang dihasilkan dapat digunakan sebagai sebagai prediksi. Model regresi semiparametrik campuran spline *truncated* dan kernel yang dihasilkan adalah sebagai berikut :

$$\begin{aligned}
\hat{y}_i = & 11,4031 - 0,1711x_{1i} - 0,01869t_{1i} - 7,8046(t_{1i} - 48,935)_+^1 + 16,3012(t_{1i} - 51,868)_+^1 + \\
& - 0,000008t_{2i} - 0,0070(t_{2i} - 10718,31)_+^1 - 0,0053(t_{2i} - 11908,23)_+^1 - 0,069t_{3i} + \\
& 0,4229(t_{3i} - 5,271)_+^1 + -2,822(t_{3i} - 6,213)_+^1 + 3,4746(t_{3i} - 7,155)_+^1 + \\
& \sum_{i=1}^n \left( \frac{\frac{1}{0,0376} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{0,0376} \right)^2}}{\sum_{i=1}^n \frac{1}{0,0376} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_1 - z_{1i}}{0,0376} \right)^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{0,2596} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{0,2596} \right)^2}}{\sum_{i=1}^n \frac{1}{0,2596} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z_2 - z_{2i}}{0,2596} \right)^2}} \right) y_i
\end{aligned}$$



## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan Analisis data yang telah dilakukan dapat ditarik kesimpulan sebagai berikut.

1. Pada penelitian ini estimasi interval untuk parameter dibangun dengan menggunakan konsep *pivotal quantity* dan estimasi interval terpendek. Estimasi interval untuk parameter regresi semiparametrik campuran spline *truncated* dan kernel terbagi menjadi dua yaitu saat  $\sigma^2$  diketahui dan  $\sigma^2$  tidak diketahui.

- i) *Pivotal quantity* saat  $\sigma^2$  diketahui adalah

$$Z_j(x_1, x_2, \dots, x_n) = \frac{\hat{\tilde{\gamma}}_j - \tilde{\gamma}_j}{\sqrt{\sigma^2 w_j}}$$

- ii) Estimasi interval saat  $\sigma^2$  diketahui adalah

$$P\left(\hat{\tilde{\gamma}}_j - v\sqrt{\sigma^2 w_{jj}} \leq \tilde{\gamma}_j \leq \hat{\tilde{\gamma}}_j + v\sqrt{\sigma^2 w_{jj}}\right) = 1 - \alpha$$

dengan  $w_{jj}$  merupakan elemen diagonal ke- $j$  dari  $(\mathbf{T}^r \mathbf{T})^{-1}$

- iii) *Pivotal quantity* saat  $\sigma^2$  tidak diketahui adalah

$$T_j(x_1, x_2, \dots, x_n) = \frac{\hat{\tilde{\gamma}}_j - \tilde{\gamma}_j}{\sqrt{MSE(\mathbf{T}^r \mathbf{T})_{jj}^{-1}}}$$

$$T_j(x_1, x_2, \dots, x_n) = \frac{\hat{\tilde{\gamma}}_j - \tilde{\gamma}_j}{\sqrt{\frac{\tilde{\mathbf{y}}^r (\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r) \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))} (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}}$$

- iv) Estimasi interval saat  $\sigma^2$  tidak diketahui adalah

$$P\left(\hat{\tilde{\gamma}}_j - t_{\frac{\alpha}{2}, (n - (1 + p + q(r+1)))} \sqrt{\frac{\tilde{\mathbf{y}}^r [\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r] \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))} (\mathbf{T}^r \mathbf{T})_{jj}^{-1}} \leq \tilde{\gamma}_j \leq \hat{\tilde{\gamma}}_j + t_{\frac{\alpha}{2}, (n - (1 + p + q(r+1)))} \sqrt{\frac{\tilde{\mathbf{y}}^r [\mathbf{I} - \mathbf{T}(\mathbf{T}^r \mathbf{T})^{-1} \mathbf{T}^r] \tilde{\mathbf{y}}}{n - (1 + p + q(r+1))} (\mathbf{T}^r \mathbf{T})_{jj}^{-1}}\right) = 1 - \alpha$$

dimana  $(\mathbf{T}^r\mathbf{T})_{jj}^{-1}$  adalah elemen diagonal ke- $j$  dari matriks  $(\mathbf{T}^r\mathbf{T})^{-1}$

2. Model regresi semiparametrik campuran spline *truncated* dan kernel terbaik yang terbentuk untuk aplikasi pada data TPT di Indonesia Tahun 2016 adalah sebagai berikut.

i) Model regresi semiparametrik campuran spline *truncated* dan kernel terbaik.

$$\hat{y}_i = 11,4031 - 0,1711x_{1i} - 0,01869t_{1i} - 7,8046(t_{1i} - 48,935)_+^1 + 16,3012(t_{1i} - 51,868)_+^1 + \\ - 0,000008t_{2i} - 0,0070(t_{2i} - 10718,31)_+^1 - 0,0053(t_{2i} - 11908,23)_+^1 - 0,069t_{3i} + \\ 0,4229(t_{3i} - 5,271)_+^1 + -2,822(t_{3i} - 6,213)_+^1 + 3,4746(t_{3i} - 7,155)_+^1 + \\ \sum_{i=1}^n \left( \frac{\frac{1}{0,0376} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z_1 - z_{1i}}{0,0376})^2}}{\sum_{i=1}^n \frac{1}{0,0376} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z_1 - z_{1i}}{0,0376})^2}} \right) y_i + \sum_{i=1}^n \left( \frac{\frac{1}{0,2596} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z_2 - z_{2i}}{0,2596})^2}}{\sum_{i=1}^n \frac{1}{0,2596} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z_2 - z_{2i}}{0,2596})^2}} \right) y_i$$

ii) Dari enam variabel bebas yang digunakan dimana satu variabel komponen parametrik dan tiga variabel komponen spline memberikan pengaruh yang signifikan terhadap model, sedangkan dua variabel lainnya didekati dengan fungsi kernel.

iii) Nilai  $R^2$  adjusted yang dihasilkan sebesar 92,53 persen, ini artinya variabel yang digunakan dapat menjelaskan model sebesar 92,53 persen, sehingga model layak digunakan.

## 5.2 Saran

1. Pada penelitian ini terbatas pada data cross section, sehingga pada penelitian berikutnya dapat dikembangkan interval konfidensi menggunakan data panel pada regresi semiparametrik.
2. Karena keterbatasan pemograman, peneliti hanya menggunakan spline linier dengan maksimal 3 titik knot. Untuk penelitian selanjutnya sebaiknya di coba menggunakan kuadratik, kubik dan lebih banyak titik knot maupun kombinasi knot yang digunakan.

3. Dalam menentukan suatu pola hubungan apakah didekati dengan parametrik maupun nonparametrik baik itu spline, kernel ataupun yang lain, sebaiknya digunakan metode lain, baik pendekatan secara matematis maupun nonmatematis sehingga lebih jelas dalam justifikasi model yang akan digunakan.

## 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.
- Alghofari, F. (2010), “Analisis Tingkat Pengangguran di Indonesia Tahun 1980-2007”, *Jurnal Pengangguran*, 1, 1: 22-52.
- Astuti, W.I., Ratnasari, V. dan Wibowo, Wahyu. (2017), “Analisis Faktor yang Berpengaruh Terhadap Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Menggunakan Regresi Data Panel”. *Jurnal Sains dan Seni ITS*, Vol 6 No.1, hal D-150-D156.
- Antoniadis, A., Gregore, G., and 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.
- BPS, (2016). *Indikator Pasar Tenaga Kerja Indonesia Agustus 2016*, BPS, Jakarta.
- Budiantara, I.N. (2005), “Regresi Spline Linier”, *Makalah Pembicara Utama pada Seminar Nasional Matematika*, FMIPA Universitas Diponegoro, Semarang.
- \_\_\_\_\_.(2009). *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang, Pidato Pengukuhan Untuk Jabatan Guru Besar pada Jurusan Statistika, Institut Teknologi Sepuluh Nopember (ITS) Surabaya*. Surabaya: ITS Press.
- Budiantara, I N. dan Mulianah. (2007), Pemilihan Bandwidth Optimal dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Sigma*, Vol. 10, No. 2, Hal. 159-166.
- 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.
- Eubank, R. (1999), *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.
- Elhorst, J.P, (2003) “The Mystery of Regional Unemployment Differentials : Theoretical and Empirical Explanations”, *Journal of Economics Surveys* Vol.17, No.5,

- Faidah, D.Y. (2012), *MODEL TOBIT Pada Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka (TPT) Perempuan*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Frenkel, R. and Ros, J. (2006). *Unemployment next term and the real exchange rate in Latin America. World Development*. 34(4), 631-646.
- Gujarati, D. N., dan Porter, D. C. (2015). *Dasar-Dasar Ekonometrika Volume 1*, 2nd Edition. Jakarta: Penerbit Salemba Empat.
- Hardle, W. (1994), *Applied Nonparametric Regression*, Springer-Verlag, Berlin.
- Hasan, M.F. (2005). “Refleksi : Masalah Pengangguran dan Kemiskinan”, *Bisnis dan Ekonomi Politik*, 6(3).
- 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.
- Idris dan Yanuar, R. (2007). *Ekonomi Sumber Daya Manusia*, FE UNP
- Intansari, I.A., (2016). “*Inferensi Statistik Untuk Kurva Regresi Nonparametrik Spline Kuadrat dan Aplikasinya Pada Data ASFR Di Provinsi Bali*”. Tesis, Jurusan Statistika, FMIPA ITS, Surabaya.
- Khusniawati, (2017), *Pengujian Hipotesis Parameter Komponen Spline dalam Model Regresi Nonparametrik Campuran Spline dan Kernel* Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Kutner, M.H., C.J. Nachtsheim., dan J. Neter. (2004). *Applied Linear Regression Models. 4th ed.* New York: McGraw-Hill Companies, Inc.
- Loklomin, B.H. (2017), *Estimasi Interval Parameter Model Spline Truncated Dalam Regresi Semiparametrik (Aplikasi : Pada Data Indeks Pembangunan Indonesia)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Mankiw, N.G. (2006). *Makroekonomi Edisi Keenam*. Jakarta: Erlangga.
- Mongomery, D. C., Peck., E. A., and Vining, G. G. (2012). *Introduction to Linear Regression Analysis Fifth Edition*, New Jersey: John Wiley & Sons.
- Mulyadi. 2003. *Ekonomi Sumber Daya Manusia Dalam Perspektif Pembangunan*. Jakarta : Raja Grafindo Persada
- Nisa, K. (2017), *Estimasi Interval Parameter Model Spline Truncated Dalam Regresi Semiparametrik (Aplikasi : Pada Data Indeks Pembangunan Indonesia)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Nafi', M., Budiantara., I.N. (2008). *Estimasi Interval Spline Dalam Regresi Nonparametrik. Tesis*. Surabaya: Institut Teknologi Sepuluh Nopember

- Purnomo, A., A., S., I. (2016), *Estimator Campuran Kernel dan Regresi Spline Truncated Linier Multivariabel dalam Regresi Nonparametrik*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Rachim, R.J. (2013), *Analisis Pengaruh Upah Minimum Provinsi, Pengeluaran Pemerintah, Investasi Swasta Dan Jumlah Penduduk Terhadap Pengangguran Terbuka Di Provinsi Sulawesi Selatan Periode 1996-2010*, Skripsi, Universitas Hasanuddin Makassar, Makassar.
- Ramadhani, R., K. (2016), *Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Pengangguran Terbuka Menggunakan Regresi Nonparametrik Spline di Indonesia*, Tugas Akhir, Institut Teknologi Sepuluh Nopember, Surabaya.
- Remi, S.,S. (2006), Korelasi Pembangunan Ekonomi, Manusia, dan Kemiskinan di Indonesia. *Bisnis & Ekonomi Politik*, 7(1).
- Rencher, A.,and Schaalje, G.B (2008), *Linear Models in Statistics 2<sup>nd</sup> Edition*, John Wiley & Sonc Inc, New Jersey.
- 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.
- Sari, R.S. dan Budiantara, I.N. (2012), “Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Pendekatan Regresi Spline Multivariabel”. *Jurnal Sains dan Seni ITS*, Vol. 1 No.1, hal D236-D241.
- Setiawan, R.N.S. (2017), *Estimasi Interval untuk Parameter Model Regresi Nonparametrik Spline Truncated Multivariabel ( Studi Kasus Pada Data Indeks Pembangunan Gender di Provinsi Jawa Timur*, Tesis, Jurusan Statistika Institut Teknologi Sepuluh Nopember, Surabaya.
- Setiawan, B.H, (2017), *Estimasi Interval Parameter Model Spline Truncated Dalam Regresi Semiparametrik (Aplikasi : Pada Data Indeks Pembangunan Indonesia)*, Tesis, Institut Teknologi Sepuluh Nopember, Surabaya.
- Sukirno, S. (2004). *Makro Ekonomi. Edisi Ketiga*. ,PT. Raja Grafindo Persada, Jakarta
- Sudiarsa, I. W., Budiantara, I. N., Suhartono, 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.
- Tantri, E. dan Ratnasari, V. (2016), “Pengaruh Indikator Kependudukan Terhadap Tingkat Pengangguran Terbuka di Indonesia dengan Pendekatan Regresi Panel”. *Jurnal Sains dan Seni ITS*, Vol 5 No.2, hal D223-D228.

Wahba, G. (1983). Bayesian Confidence Interval for the Cross Validated Smoothing Parameter in the Generalized Spline Smoothing Problems. *The Annals of Statistics*, 13, 1378-1402.

\_\_\_\_\_. (1990), *Spline Models for Observational Data*, SIAM, Pennsylvania.

Wang, Y. (1998). Spline Smoothing Models with Correlated Errors. *Journal of the American Statistical Association*, 93, 341-348.

## LAMPIRAN

Lampiran 1. Data Tingkat Pengangguran Terbuka di Indonesia Tahun 2016 dan Faktor-Faktor yang diduga Berpengaruh

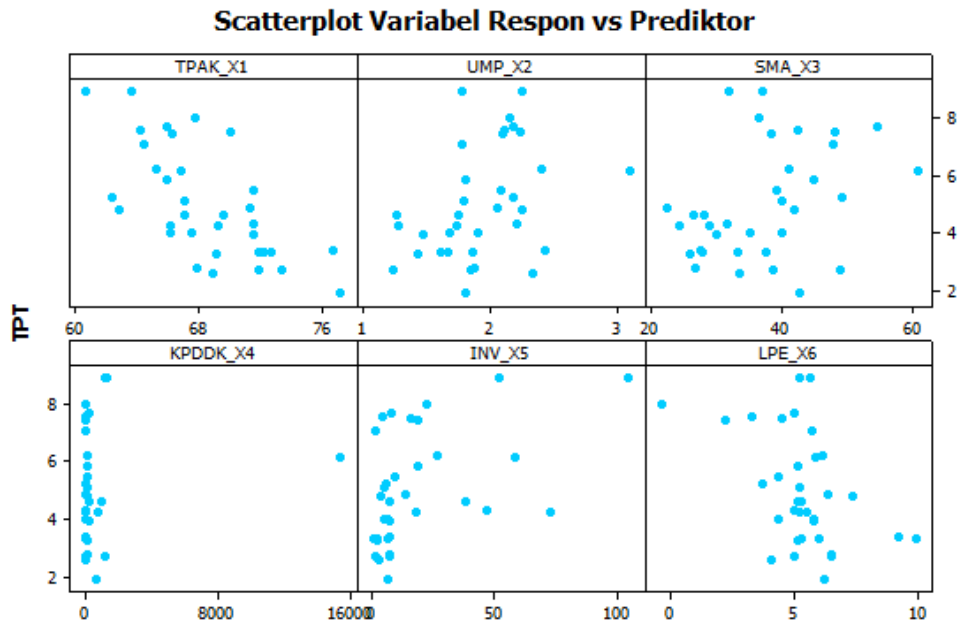
| Provinsi                  | y    | x <sub>1</sub> | x <sub>2</sub> | x <sub>3</sub> | x <sub>4</sub> | x <sub>5</sub> | x <sub>6</sub> |
|---------------------------|------|----------------|----------------|----------------|----------------|----------------|----------------|
| Aceh                      | 7.57 | 64.26          | 2.12           | 42.36          | 88             | 4.28           | 3.31           |
| Sumatera Utara            | 5.84 | 65.99          | 1.81           | 44.94          | 193            | 18.70          | 5.18           |
| Sumatera Barat            | 5.09 | 67.08          | 1.8            | 40.07          | 125            | 4.88           | 5.26           |
| Riau                      | 7.43 | 66.25          | 2.1            | 38.36          | 75             | 18.53          | 2.23           |
| Jambi                     | 4    | 67.54          | 1.91           | 35.18          | 69             | 4.72           | 4.37           |
| Sumatera Selatan          | 4.31 | 71.59          | 2.21           | 31.59          | 89             | 47.12          | 5.03           |
| Bengkulu                  | 3.3  | 72.69          | 1.61           | 37.6           | 96             | 1.71           | 5.3            |
| Lampung                   | 4.62 | 69.61          | 1.76           | 28.27          | 237            | 7.21           | 5.15           |
| Kepulauan Bangka Belitung | 2.6  | 68.93          | 2.34           | 33.62          | 85             | 2.93           | 4.11           |
| Kepulauan Riau            | 7.69 | 65.93          | 2.18           | 54.71          | 247            | 7.57           | 5.03           |
| DKI Jakarta               | 6.12 | 66.91          | 3.1            | 60.67          | 15478          | 58.72          | 5.85           |
| Jawa Barat                | 8.89 | 60.65          | 2.25           | 31.92          | 1339           | 105.34         | 5.67           |
| Jawa Tengah               | 4.63 | 67.15          | 1.27           | 26.62          | 1037           | 38.15          | 5.28           |
| DI Yogyakarta             | 2.72 | 71.96          | 1.24           | 49.01          | 1188           | 1.22           | 5.05           |
| Jawa Timur                | 4.21 | 66.14          | 1.28           | 28.84          | 817            | 72.90          | 5.55           |
| Banten                    | 8.92 | 63.66          | 1.78           | 37.14          | 1263           | 52.40          | 5.26           |
| Bali                      | 1.89 | 77.24          | 1.81           | 42.72          | 717            | 6.67           | 6.24           |
| Nusa Tenggara Barat       | 3.94 | 71.57          | 1.48           | 29.98          | 264            | 7.39           | 5.82           |
| Nusa Tenggara Timur       | 3.25 | 69.18          | 1.43           | 26.09          | 107            | 1.62           | 5.18           |
| Kalimantan Barat          | 4.23 | 69.32          | 1.74           | 24.33          | 33             | 17.62          | 5.22           |
| Kalimantan Tengah         | 4.82 | 71.3           | 2.06           | 22.53          | 17             | 13.74          | 6.36           |
| Kalimantan Selatan        | 5.45 | 71.57          | 2.09           | 39.26          | 105            | 9.56           | 4.38           |
| Kalimantan Timur          | 7.95 | 67.79          | 2.16           | 36.56          | 27             | 22.41          | -0.38          |
| Kalimantan Utara          | 5.23 | 62.4           | 2.18           | 49.19          | 9              | 5.53           | 3.75           |
| Sulawesi Utara            | 6.18 | 65.22          | 2.4            | 41.16          | 176            | 26.95          | 6.17           |
| Sulawesi Tengah           | 3.29 | 72.28          | 1.67           | 33.16          | 47             | 6.17           | 9.98           |
| Sulawesi Selatan          | 4.8  | 62.92          | 2.25           | 41.88          | 184            | 3.62           | 7.41           |
| Sulawesi Tenggara         | 2.72 | 73.47          | 1.85           | 38.74          | 67             | 6.98           | 6.51           |
| Gorontalo                 | 2.76 | 67.89          | 1.88           | 26.86          | 102            | 7.30           | 6.52           |
| Sulawesi Barat            | 3.33 | 71.9           | 1.86           | 27.98          | 78             | 0.26           | 6.03           |
| Maluku                    | 7.05 | 64.51          | 1.78           | 47.97          | 37             | 1.42           | 5.76           |
| Maluku Utara              | 4.01 | 66.19          | 1.68           | 39.98          | 37             | 5.98           | 5.77           |
| Papua Barat               | 7.46 | 70.05          | 2.24           | 48.05          | 9              | 15.96          | 4.52           |
| Papua                     | 3.35 | 76.7           | 2.44           | 27.48          | 10             | 7.33           | 9.21           |



Keterangan :

- $y$  : TPT (Tingkat Pengangguran Terbuka)
- $x_1$  : Tingkat Partisipasi Angkatan Kerja
- $x_2$  : Upah Minimum Provinsi
- $x_3$  : Persentase penduduk usia 15+ yang lulus SMA/SMK
- $x_4$  : Kepadatan Penduduk
- $x_5$  : Investasi
- $x_6$  : Laju Pertumbuhan Ekonomi

Lampiran 2. *Scatter Plot* Variabel Respon terhadap Variabel Prediktor



### Lampiran 3. Sintax R Model Semiparametric Campuran Spline Truncated dan Kernel

```
SplineKernel=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
  }
}
```

Lampiran 3. Sintax R Model Semiparametric Campuran Spline Truncated dan Kernel (Lanjutan)

```

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)))

```

### Lampiran 3. Sintax R Model Semiparametric Campuran Spline Truncated dan Kernel (Lanjutan)

```

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
  }
}

```

### Lampiran 3. Sintax R Model Semiparametric Campuran Spline Truncated dan Kernel (Lanjutan)

```

    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

```

### Lampiran 3. Syntax R Model Semiparametric Campuran Spline Truncated dan Kernel (Lanjutan)

```

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

#interval confidence
SEbeta=sqrt(diag(matMSE[i]*(pinv(t(x)%%(x))))))
SEteta=sqrt(diag(matMSE[i]*(pinv(t(Xnon)%%(Xnon))))))
tthit=pt(((1-alpha)/2),(n-length(beta[i,])-length(teta[i,])))
ba_beta[i,]=beta[i,]+tthit*SEbeta
bb_beta[i,]=beta[i,]-tthit*SEbeta
ba_teta[i,]=teta[i,]+tthit*SEteta
bb_teta[i,]=teta[i,]-tthit*SEteta
}
dataprint =
data.frame(cbind(matGCV,matRsq,matMSE,beta,bb_beta,ba_beta,
                 teta,bb_teta,ba_teta,knot,matalfa))

mbeta=ncol(beta)
mteta=ncol(teta)
mknot=ncol(knot)
malfa=ncol(matalfa)
a<-c("GCV", "Rsq", "MSE", paste("beta", c(0:(mbeta-
1))), paste("bb_beta", c(0:(mbeta-1))), paste("ba_beta", c(0:(mbeta-1))),
paste("teta", c(1:(mteta))), paste("bb_teta", c(1:(mteta))), paste("ba_teta",
,c(1:(mteta))), paste("knot", c(1:(mknot))), paste("alfa", c(1:(malfa))))
colnames(dataprint)<-a
dataprint = dataprint[order(dataprint$GCV,dataprint$MSE),]
cat("\n", "Result : ", "\n")
nprint=min(20,nrow(dataprint))
print(dataprint[1:nprint,])
res=matrix(res[,as.numeric(rownames(dataprint)[1])],ncol=1)
cat("\n", "residual : ", "\n")
print (res)
yhat1=yhat1[,as.numeric(rownames(dataprint)[1])]
write.csv(yhat1,paste('result yhat knot ', tknot, '.csv'))
write.csv(res,paste('result residual knot ', tknot, '.csv'))
write.csv(dataprint,paste('result estimation parameter knot ',
tknot, '.csv'))
list(hasil=dataprint)
}

```

## BIOGRAFI PENULIS



Penulis dilahirkan di Jepara, Jawa Tengah pada tanggal 23 Februari 1985, merupakan putra kedua dari lima bersaudara, buah cinta dari pasangan Bapak Abdul Kadir dan Ibu Murdiah. Saat ini penulis sudah berkeluarga dengan istri bernama Middia Martanti Dewi dan telah dikaruniai dua putra yang sholeh bernama Rasya Muhammad Athaya dan Fadlan Muhammad Hamizan. Penulis memulai pendidikan formalnya dari SD Negeri Panggang 1 Jepara (1991-1997), SMP Negeri 1 Jepara (1997-2000), SMU Negeri 1 Jepara (2000-2003). Penulis kemudian melanjutkan pendidikan ke jenjang sarjana di Sekolah Tinggi Ilmu Statistik (STIS) Jakarta (2003-2008) jurusan Statistik Sosial Kependudukan. Setelah menyelesaikan pendidikan DIV di STIS, penulis ditugaskan bekerja di BPS Kabupaten Morowali Provinsi Sulawesi Tengah (2008-2014), BPS Provinsi Jawa Tengah (2014-Sekarang). Pada tahun 2016 penulis memperoleh kesempatan beasiswa dari BPS untuk melanjutkan pendidikan ke jenjang S2 di Jurusan Statistika Fakultas Matematika, Komputasi dan Sains Data (FMKSD) Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Pembaca yang ingin memberikan kritik, saran dan pertanyaan mengenai penelitian ini bisa menghubungi penulis melalui email : [masfen45@gmail.com](mailto:masfen45@gmail.com).