TESIS-KS185411

PERBANDINGAN MODEL REGRESI SEMIPARAMETRIK BIRESPON LINIER DAN KUADRATIK MENGGUNAKAN SPLINE TRUNCATED

(Aplikasi Pada Data Angka Harapan Hidup dan Pengeluaran Perkapita)

ADAWIYAH ASTI KHALIL
06211850010002

Dosen Pembimbing
Prof. Dr. Drs. I Nyoman Budiantara, M.Si
Dr. Ismaini Zain, M.Si

Program Magister
Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
Surabaya
2020

TESIS – KS185411

PERBANDINGAN MODEL REGRESI SEMIPARAMETRIK BIRESPON LINIER DAN KUADRATIK MENGGUNAKAN SPLINE TRUNCATED

(Aplikasi Pada Data Angka Harapan Hidup dan Pengeluaran Perkapita)

ADAWIYAH ASTI KHALIL
06211850010002

Dosen Pembimbing
Prof. Dr. Drs. I Nyoman Budiantara, M.Si
Dr. Ismaini Zain, M.Si

Program Magister
Departemen Statistika
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember
Surabaya
2020

THESIS – KS185411

COMPARISON OF LINEAR AND QUADRATIC BI-RESPONSE SEMIPARAMETRIC REGRESSION MODELS USING SPLINE TRUNCATED

(Application to Life Expectancy Data and Per-capita Expenditures)

ADAWIYAH ASTI KHALIL
06211850010002

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

Magister Program
Department of Statistics
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember
Surabaya
2020

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Statistika (M.Stat)
di
Institut Teknologi Sepuluh Nopember

Oleh:

ADAWIYAH ASTI KHALIL
NRP: 06211850010002

Tanggal Ujian : 28 Juli 2020
Periode Wisuda : September 2020

Disetujui oleh:

Pembimbing:

1. Prof. Dr. I Nyoman Budiantara, M.Si.
NIP: 19650603 198903 1 003 ....

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

Penguji :

1. Dr. Vita Ratnasari, M.Si.
NIP: 19700910 199702 2 001 ....

2. Jerry Dwi Trijoyo Purnomo, M.Si, Ph.D.
NIP: 19810223 200812 1 003 ....

Kepala Departemen Statistika
Fakultas Sains dan Analitika Data

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
FAKULTAS SAINS DAN ANALITIKA DATA
DEPARTEMEN STATISTIKA

Dr. Kartika Fithriasari, M.Si.
NIP: 19691212 199303 2 002

(Halaman ini sengaja dikosongkan)

PERBANDINGAN MODEL REGRESI SEMIPARAMETRIK BIRESPON LINIER DAN KUADRATIK MENGGUNAKAN SPLINE TRUNCATED (Aplikasi Pada Data Angka Harapan Hidup dan Pengeluaran Perkapita)

Nama Mahasiswa	: Adawiyah Asti Khalil
NRP	: 06211850010002
Pembimbing	: Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
	Dr. Ismaini Zain, M.Si.

ABSTRAK

Regresi semiparametrik birespon merupakan model regresi yang terdiri dari dua variabel respon dan memiliki dua komponen yaitu komponen parametrik yang pola antara variabel prediktor dan variabel respon diketahui dapat berpola linier, kuadratik dan komponen nonparametrik memiliki pola antara variabel prediktor dan variabel respon tidak diketahui. Pada penelitian ini data yang digunakan berupa data Angka Harapan Hidup dan data pengeluaran perkapita. Variabel prediktor yang digunakan yaitu rata-rata lama sekolah, tingkat pengangguran terbuka, dan persentase penduduk miskin. Penelitian ini memiliki dua tujuan, pertama mengkaji estimasi parameter untuk regresi semiparametrik birespon *spline truncated* dengan komponen parametrik kuadratik menggunakan metode *Weighted Least Square* (WLS). Tujuan kedua pada penelitian ini adalah untuk memodelkan data angka harapan hidup dan pengeluaran perkapita menggunakan model regresi semiparametrik birespon komponen parametrik linier dan kuadratik pendekatan *spline truncated* linier dengan menggunakan metode *Generalized Cross Validation* (GCV), dengan memilih titik-titik knot optimal berdasarkan nilai GCV terkecil. Estimasi model regresi semiparametrik birespon *spline truncated* komponen parametrik kuadratik memiliki perbedaan dengan komponen parametrik linier yaitu pada struktur matriks $\mathbf{X}$ (variabel prediktor). Model regresi semiparametrik birespon dengan komponen parametrik kuadratik menghasilkan nilai GCV yang lebih kecil dan nilai koefisien determinasi yang lebih besar jika dibandingkan dengan komponen parametrik linier. Nilai GCV untuk komponen parametrik linier dan kuadratik adalah sebesar 1.605 dan 1.493. Dengan nilai koefisien determinasi untuk komponen parametrik linier dan kuadratik masing-masing sebesar 99.85% dan 99.87%.

Kata kunci: Birespon, Semiparametrik, *Spline Truncated, Weighted Least Square*

(Halaman ini sengaja dikosongkan)

COMPARISON OF LINEAR AND QUADRATIC BI-RESPONSE SEMIPARAMETRIC REGRESSION MODELS USING SPLINE TRUNCATED
(Application to Life Expectancy Data and Per-capita Expenditures)

Name of Student	: Adawiyah Asti Khalil
Student Identity Number	: 06211850010002
Supervisors	: Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
	Dr. Ismaini Zain, M.Si

ABSTRACT

Bi-response semiparametric regression is a regression model consists of two response variables and has two components, namely the parametric component that has a pattern between predictor variables and response variables are known to be linear or quadratic in the pattern and the nonparametric components has a pattern between the predictor variable and the response variable is unknown. In this study the data used is life expectancy data and per capita expenditure data. The predictor variables used are the average length of schooling, the open unemployment rate, and the percentage of the poor population. This study has two objectives, the first of which is to assess the estimation of the parameter for a bi-response semiparametric regression spline truncated model with quadratic parametric component using Weighted Least Square (WLS) method. The second objective in this study is to do a model life expectancy and expenditure data per capita using a linear and quadratic components bi-response semiparametric regression model with linear Spline Truncated approach using the Generalized Cross-Validation (GCV) method, by selecting optimal knot points based on the smallest GCV values. The estimation semiparametric regression model of bi-response spline truncated quadratic parametric component has a difference with the linear parametric component that is in the $\mathbf{X}$ matrix structure (predictor variabel matrix). Bi-response semiparametric regression model with quadratic parametric component produce smaller GCV values and greater coefficient of determination values when compared to linear parametric components. The GCV values for linear parametric and quadratic parametric components are 1.605 and 1.493. With the coefficient of determination for the linear and quadratic parametric components are 99.85% and 99.87% respectively.

Keywoards: Bi-response, Semiparametric, Spline Truncated, Weighted Least Square

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, serta shalawat dan salam selalu tercurahkan pada Nabi Muhammad SAW suri tauladan terbaik dalam kehidupan ini, sehingga penulis dapat menyelesaikan penyusunan Tesis yang berjudul "**Perbandingan Model Regresi Semiparametrik Birespon Linier Dan Kuadratik Menggunakan Spline Truncated**" (Aplikasi Pada Data Angka Harapan Hidup dan Pengeluaran Perkapita). Terselesaikannya penyusunan Tesis ini tidak lepas dari bantuan berbagai pihak, baik bantuan secara langsung maupun tidak langsung, sehingga pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Kartika Fithriasari, M.Si. selaku Kepala Departemen Statistika Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember, Surabaya.
2. Ibu Santi Wulan Purnami, M.Si., Ph.D. selaku Sekertaris Departemen I (Bidang Akademik, Kemahasiswaan, Penelitian, dan Pengabdian Kepada Masyarakat).
3. Ibu Dr. Vita Ratnasari, M.Si. selaku Sekertaris Departemen II (Bidang Sumber Daya Keuangan, Sumber Daya Manusia, dan Sarana Prasarana) sekaligus dosen penguji tesis yang telah memberikan banyak masukan dan saran kepada penulis untuk penyusunan tesis yang lebih baik.
4. Bapak Dr.rer.pol., Dedy Dwi Prasetyo, S.Si., M.Si. selaku Kepala Program Studi Pascasarjana Statistika, Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember, Surabaya.
5. Prof. Dr. Drs. I Nyoman Budiantara, M.Si. dan Ibu Dr. Ismaini Zain, M.Si selaku dosen pembimbing tesis yang telah meluangkan banyak waktu untuk memberikan bimbingan ilmu, arahan, serta motivasi kepada penulis dalam penyusunan tesis ini.
6. Bapak Jerry Dwi Trijoyo Purnomo, M.Si, Ph.D. selaku dosen penguji tesis yang telah memberikan banyak masukan dan saran kepada penulis untuk penyusunan tesis yang lebih baik.
7. Bapak Dr. Drs. Agus Suharsono, M.S. selaku dosen wali yang telah memberikan motivasi dan masukan demi kelancaran terselesaikannya studi.

8. Bapak dan Ibu dosen Departemen Statistika, FSAD, ITS yang telah memberikan banyak sekali bekal ilmu, motivasi, serta arahan kepada penulis selama masa perkuliahan.
9. Bapak dan Ibu staff Departemen Statistika, FSAD, ITS yang telah memfasilitasi dan membantu penulis selama menempuh pendidikan.
10. Kedua orang tua tercinta, Bapak Abaad dan Ibu Suhati, adik tersayang Adelia Nursamawa Khalil, Anggi Sukma Putri Khalil, seluruh keluarga besar yang di Lombok dan Surabaya, terima kasih yang sebesar-besarnya atas segala dukungan berupa doa, semangat, moral, dan materiil yang sangat berarti hingga akhir penyusunan tesis.
11. Sahabat-sahabat: Anjar Wicitra Wening, Ramli, mb Lulu, Nimas, Nazmi, Nisa, dan Istin yang senantiasa menemani diskusi, berbagi ilmu, memberikan semangat dan doa dalam penyelesaian tesis ini.
12. Teman seperjuangan tesis, Rossy, Kinanti, Almira, Yusril, mb Aisah, mb Dini, Ika, Vania, Sania, Ria yang telah saling menyemangati dan mendoakan.
13. Teman-teman S2 Statistika 2018 Ganjil, S2 Statistika 2017, HIMMPAS ITS, atas segala bantuan, doa, dan persaudaraan semasa perkuliahan.
14. Seluruh pihak yang telah banyak membantu penulis dalam penyusunan tesis dan belum bisa untuk penulis sebutkan satu per satu. Semoga Allah SWT membalas semua kebaikan dengan kebaikan yang lebih baik.

Penulis menyadari bahwa dalam penulisan tesis ini masih terdapat kekurangan baik dari segi penyusunan, bahasa atau penulisan, maka penulis mengharapkan kritik dan saran yang bersifat membangun, sebagai bekal pengalaman bagi penulis untuk lebih baik di masa yang akan datang. Semoga tesis ini dapat bermanfaat bagi pembaca.

Surabaya, Agustus 2020

Adawiyah Asti Khalil

DAFTAR ISI

DAFTAR GAMBAR

Halaman

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Halaman

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Halaman

(Halaman ini sengaja dikosongkan)

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Salah satu metode statistika yang dapat digunakan untuk mengestimasi hubungan antara variabel prediktor dan variabel respon adalah analisis regresi. Analisis regresi mendeskripsikan fungsi yang disebut dengan kurva regresi (Gujarati dan Porter, 2015). Pendekatan parametrik digunakan apabila kurva regresi antara variabel prediktor dan variabel respon diasumsikan mengikuti pola yang bentuknya diketahui contohnya mengikuti pola linier, kuadratik, dan lainnya. Pendekatan nonparametrik digunakan apabila kurva regresi antara variabel prediktor dan variabel respon diasumsikan bentuknya tidak diketahui atau tidak mengikuti pola tertentu (Eubank, 1999). Pendekatan semiparametrik merupakan kombinasi dari pendekatan parametrik dan pendekatan nonparametrik yaitu memiliki bentuk kurva regresi yang diasumsikan sebagian mengikuti pola terntentu dan sebagian tidak mengikuti pola tertentu atau tidak diketahui (Hardle, 1990).

Beberapa metode yang banyak digunakan untuk mengestimasi kurva regresi nonparametrik atau regresi semiparametrik, diantaranya Spline, Kernel, dan Deret Fourier, wavelets, (MARS) *Multivariate Adaptive Regression Spline* (Budiantara, 2009). Beberapa penelitian sebelumnya yang menggunakan metode tersebut diantaranya Mardianto (2015) menggunakan estimator Deret Fourier, Chamidah dan Afifah (2017) menggunakan estimator Kernel, serta Pratiwi (2017) yang menggunakan estimator Spline.

Model regresi nonparametrik dan semiparametrik yang banyak digunakan dalam beberapa kurun waktu terakhir adalah regresi spline. Metode spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu dan dapat digunakan untuk mengatasi atau mengurangi pola data yang mengalami peningkatan tajam dengan bantuan titik knot (Budiantara, 2014). Spline merupakan model yang memiliki fleksibilitas yang tinggi dan mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik (Eubank, 1999). Pada kenyataannya, dalam kasus sehari-hari terdapat kasus yang menggunakan analisis regresi lebih dari satu respon.

Misalkan terdapat korelasi antara variabel respon tersebut, jika dilakukan analisis secara parsial pada model regresi dua respon tersebut maka model yang dihasilkan bukanlah model yang optimal sehingga para statistikawan mengembangkan model regresi birespon. Model regresi yang memiliki lebih dari satu variabel respon yang saling berkorelasi dengan satu atau lebih variabel prediktor disebut model regresi birespon (Rencher, 2008).

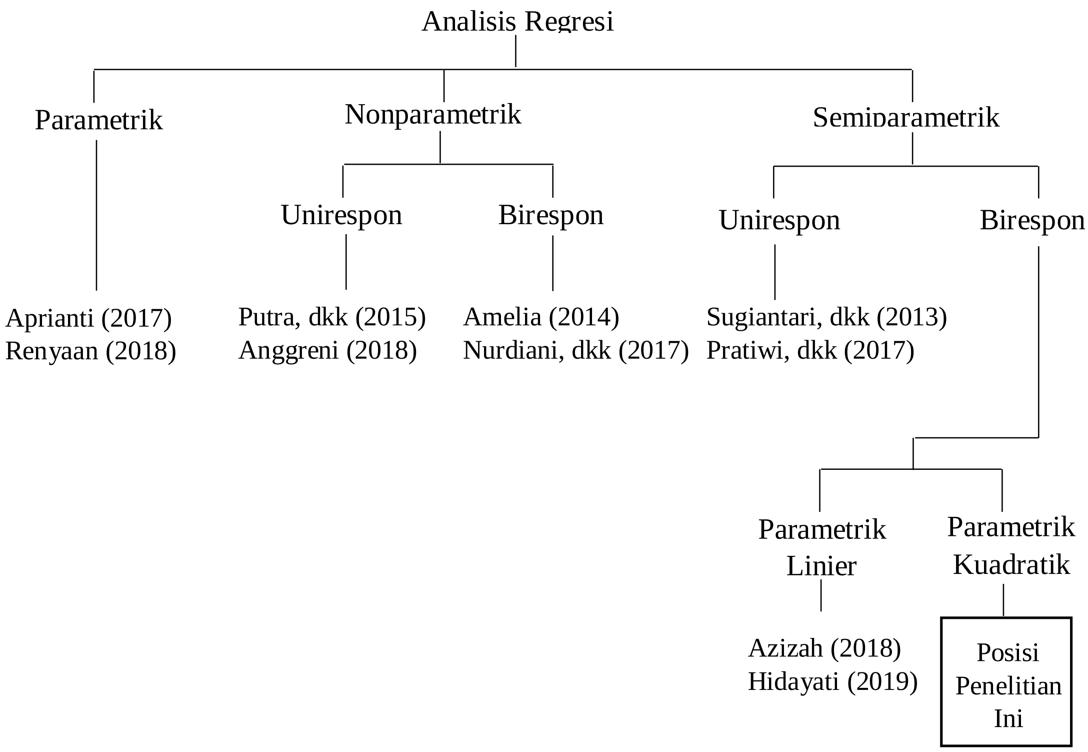
Metode spline yang banyak digunakan oleh beberapa peneliti adalah metode spline truncated. Beberapa penelitian sebelumnya mengenai metode spline truncated untuk mengetahui model regresi nonparametrik dan semiparametrik telah dilakukan Pratiwi (2017) dan Prawanti (2019). Selanjutnya, penelitian dengan menggunakan metode spline truncated dengan menggunakan data birespon telah dilakukan oleh Amelia (2014) dan Nurdiani (2017) mengenai model regresi nonparametrik birespon, sedangkan Azizah (2018), mengenai regresi semiparametrik birespon. Pada penelitian-penelitian tersebut, variabel respon yang digunakan dua atau birespon dan berfokus pada pencarian kurva regresi nonparametrik atau semiparametrik untuk komponen parametrik linier. Pada kasus tertentu, apabila komponen parametrik tidak dapat didekati dengan pola linier atau komponen parametrik tidak mengikuti pola tertentu, maka komponen parametrik tersebut dapat didekati dengan pola kuadratik. Salah satu studi kasus dengan dua variabel respon yaitu komponen penyusun Indeks Pembangunan Manusia (IPM), dimana dua di antara komponen penyusun IPM juga termasuk dalam Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals*/SDGs).

Pada tahun 2015, Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals*/SDGs) telah disahkan oleh Perserikatan Bangsa Bangsa (PBB) sebagai agenda pembangunan global yang baru untuk periode 2016-2030. Konsep SDGs ini berkaitan dengan beberapa isu krusial global diantaranya, perubahan iklim, perlindungan sosial, pembangunan, ketahanan pangan dan energi. SDGs dibentuk dengan 17 tujuan yang hendak dicapai. Di antara 17 tujuan SDGs, ada beberapa tujuan yang berhubungan dengan Indeks Pembangunan Manusia (IPM), salah satunya yaitu tujuan ketiga, meningkatkan kesejahteraan dan memastikan kehidupan yang sehat untuk penduduk segala usia (BPS, 2019). Terdapat beberapa indikator yang dapat menjelaskan hasil pembangunan pada dimensi kesehatan dan dimensi kesejahteraan, antara lain angka harapan hidup dan pendapatan riil per kapita yang disesuaikan dengan daya beli (Todaro dan Smith, 2012). Proksi umur panjang dan sehat yang digunakan dalam pembangunan adalah indikator angka harapan hidup saat lahir. Indikator ini menjadi salah satu indikator gambaran kesehatan masyarakat, sedangkan indikator yang umum digunakan dalam menggambarkan tingkat kesejahteraan adalah pengeluaran perkapita (BPS, 2019).

Data terbaru menunjukkan bahwa capaian indeks kesehatan dan indeks standar hidup layak di Indonesia pada tahun 2018 masing-masing mencapai 0,78 dan 0,73. Untuk provinsi Jawa Timur indeks kesehatan dan standar hidup layak masing-masing 0,78 dan 0,74. Untuk kab/kota yang memiliki indeks kesehatan tertinggi yaitu Surabaya, Sidoarjo, dan Kediri sebesar 0,83 dan terendah Probolingo dan Bondowoso sebesar 0,71. Sebanyak 17 kabupaten/kota yang indeks kesehatan di atas rata-rata Jawa Timur (Kuadran I), umumnya daerah perkotaan. Sedangkan yang tercatat di Kuadran III atau mempunyai indeks kesehatan yang lebih rendah daripada Jawa Timur sebanyak 11 kabupaten/kota. Untuk indeks standar hidup layak kota Surabaya memiliki indeks pengeluaran tertinggi sebesar 0,87 dan terendah sebesar 0,65 yaitu kabupaten Pacitan, Bangkalan, dan Sampang. Terdapat 14 kabupaten/kota yang mempunyai indeks pengeluaran lebih tinggi dari pada Jawa Timur (Kuadran I), sedangkan 18 kabupaten/kota lainnya masuk di Kuadran III atau mempunyai indeks pengeluaran di bawah Jawa Timur (BPS Jatim, 2019).

Beberapa penelitian mengenai indeks kesehatan maupun terhadap indikator penyusunnya telah dilakukan, diantaranya Sugiantari dan Budiantara (2013) melakukan penelitian terhadap faktor-faktor yang mempengaruhi angka harapan hidup di Jawa Timur menggunakan regresi semiparametrik spline. Maully (2014) meneliti faktor-faktor yang mempengaruhi indeks kesehatan kabupaten/kota di Jawa Timur dengan menggunakan regresi logistik. Ardianti, Wibisono, dan Jumiati (2015) melakukan pemodelan angka harapan hidup di Kabupaten Jember dengan metode analisis regresi berganda. Demikian juga penelitian tentang indeks pengeluaran maupun terhadap indikator penyusunnya telah dilakukan. Pertiwi (2012) melakukan pemodelan pengeluaran per kapita per kabupaten/kota di Kalimantan Barat dengan menggunakan metode hirarki bayesian. Fitriani (2015) melakukan analisis faktor-faktor yang mempengaruhi daya beli (pengeluaran per kapita) masyarakat di Jawa Barat dengan menggunakan analisis regresi.

Merujuk pada penelitian-penelitian sebelumnya maka pada penelitian ini akan dikaji bentuk estimator kurva regresi birespon semiparametrik spline *truncated* dengan komponen parametrik kuadratik dan selanjutnya diaplikasikan pada data angka harapan hidup dan pengeluaran perkapita di provinsi Jawa Timur.



Gambar 1.1 Posisi Penelitian Ini Dibandingkan Penelitian Sebelumnya

1.2 Rumusan Masalah

Berdasarkan latar belakang yang diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana menentukan estimasi kurva regresi semiparametrik birespon kuadratik menggunakan Spline Truncated?
2. Bagaimana aplikasi regresi semiparametrik birespon komponen parametrik kuadratik jika dibandingkan dengan birespon komponen parametrik linier menggunakan Spline Truncated pada data angka harapan hidup dan pengeluaran perkapita di provinsi Jawa Timur?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendapatkan estimator kurva regresi semiparametrik birespon komponen parametrik linier dan komponen parametrik kuadratik menggunakan pendekatan Spline Truncated.
2. Membandingkan model regresi semiparametrik birespon komponen parametrik linier dan komponen parametrik kuadratik dengan pendekatan Spline Truncated pada data angka harapan hidup dan pengeluaran perkapita di Jawa Timur.

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Menambah wawasan keilmuan dan pengetahuan mengenai model regresi semiparametrik birespon menggunakan pendekatan Spline Truncated.
2. Menambah pemahaman aplikasi model regresi semiparametrik birespon pendekatan Spline Truncated pada data umur harapan hidup dan pengeluaran perkapita.

1.5 Batasan Masalah

Berdasarkan tujuan permasalahan yang telah diuraiakan sebelumnya, maka batasan masalah pada penelitian ini adalah sebagai berikut.

1. Estimasi model regresi semiparametrik birespon pendekatan Spline *Truncated* menggunakan *Weighted Least Square* (WLS).
2. Pemilihan model titik knot optimal menggunakan metode *Generalized Cross Validation* (GCV) dengan menggunakan satu titik knot.
3. Fungsi linier dan fungsi kuadratik yang dimaksud adalah data pada komponen parametrik.
4. Variabel prediktor satu mempengaruhi variabel respon satu dan respon dua dengan pola parametrik, sedangkan variabel prediktor dua mempengaruhi variabel respon satu dan respon dua dengan pola nonparametrik.
5. Fungsi Spline *Truncated* yang digunakan adalah fungsi Spline *Truncated* linier

(Halaman ini sengaja dikosongkan)

BAB 2

TINJAUAN PUSTAKA

2.1 Analisis Regresi Parametrik, Nonparametrik, dan Semiparametrik

Analisis regresi adalah analisis statistika yang digunakan untuk menyelidiki hubungan antara variabel prediktor dengan variabel respon melalui etimasi kurva regresi (Drapper dan Smith, 1992). Estimasi kurva regresi dapat dilakukan dengan tiga pendekatan, yaitu pendekatan regresi parametrik, pendekatan regresi nonparametrik, dan pendekatan regresi semiparametrik (Eubank, 1999).

Regresi parametrik digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor dengan asumsi bentuk estimasi kurva regresi diketahui. Menurut Ruppert, dkk. (2003) selain memiliki asumsi kurva estimasi diketahui, regresi parametrik juga memiliki asumsi residual bersifat identik, independen, dan berdistribusi normal $\varepsilon_i \sim IIDN\ (0,\sigma^2)$. Salah satu bentuk regresi parametrik dapat dinyatakan sebagai model regresi linier berganda, secara umum dapat dituliskan dalam notasi matrik sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_i x_{i1} + e_i \quad ; i=1,2,\ldots n \tag{2.1}$$

Dengan n adalah banyaknya observasi, sedangkan y merupakan variabel respon, x merupakan variabel prediktor sebanyak D, dan e_i residual yang berdistribusi $\varepsilon_i \sim IIDN(0,\sigma^2)$. Persamaan (2.1) dapat ditulis dalam bentuk matriks berukuran n x l sebagai berikut.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

dengan

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{D1} \\ 1 & x_{12} & \cdots & x_{D2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{Dn} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_D \end{pmatrix}, \text{dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Regresi nonparametrik adalah suatu pendekatan analisis regresi yang digunakan apabila estimasi kurva regresi diasumsikan tidak memiliki pola atau bentuk tertentu (Eubank, 1999). Kurva regresi dalam regresi nonparametrik diasumsikan halus (*smooth*), sehingga memiliki fleksibilitas yang tinggi karena bentuk estimasi kurva regresi dapat terbentuk mengikuti data tanpa dipengaruhi oleh faktor subyektivitas peneliti (Budiantara, 2009). Diberikan data (x_i, y_i) dengan hubungan antara x_i dan y_i diasumsikan mengikuti model regresi nonparametrik dengan rumus umum sebagai berikut (Eubank, 1999).

$$\boldsymbol{y} = \boldsymbol{T\alpha} + \boldsymbol{\varepsilon} \tag{2.3}$$

dengan $\boldsymbol{y}$ merupakan variabel respon, dengan T merupakan matriks yang terdiri dari variabel prediktor untuk komponen nonparametrik dan titik knot, dan ε adalah vektor error yang berdistribusi normal independen dengan mean 0 dan σ^2.

Regresi semiparametrik merupakan gabungan dari regresi parametrik dan regresi nonparametrik. Pendekatan regresi semiparametrik dilakukan apabila sebagian estimasi kurva regresi diasumsikan polanya diketahui dan sebagian polanya tidak diketahui. Bentuk regresi semiparametrik dapat ditulis sebagai berikut.

$$y_i = f(x_i) + g(t_i) + \varepsilon_i, \quad i=1,2,\ldots,n \tag{2.4}$$

dengan y_i adalah variabel respon untuk pengamatan ke-i, sedangkan $f(x_i)$ merupakan fungsi regresi komponen parametrik pada pengamatan ke-i, $g(t_i)$ fungsi regresi komponen nonparametrik pada pengamatan ke-i, dan ε_i adalah error random $\varepsilon_i \sim N(0, \sigma^2)$. Sehingga, berdasarkan persamaan 2.2 dan persamaan 2.3 maka persamaan 2.4 dapat ditulis dalam bentuk matriks berikut.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{T\alpha} + \boldsymbol{\varepsilon} \tag{2.5}$$

dengan $\boldsymbol{y}$, β, $\boldsymbol{\alpha}$, dan ε merupakan vektor sedangkan $\boldsymbol{X}$ dan $\boldsymbol{T}$ adalah matriks.

2.2 Spline Truncated

Spline merupakan salah satu metode estimasi untuk model regresi nonparametrik dan semiparametrik yang sering digunakan. Metode spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu dan dapat digunakan untuk mengatasi atau mengurangi pola data yang mengalami peningkatan tajam dengan bantuan titik knot (Budiantara, 2014). Spline merupakan model yang memiliki fleksibilitas yang tinggi dan mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik (Eubank, 1999). Fungsi spline berderajat *p* dengan titik-titik knot $K_1, K_2, \ldots, K$ adalah sebagai berikut (Hardle, 1990):

$$g(t_i) = \alpha_0 + \sum_{p=1}^{P} \alpha_p t_i^{\,j} + \sum_{k=1}^{K} \alpha_{p+k} (t_i - K_k)_+^p \qquad (2.6)$$

dengan fungsi *truncated* sebagai berikut:

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

dimana α_p dengan $p = 1, 2, \ldots, P$ dan α_{p+k} dengan $k = 1, 2, \ldots, K$ merupakan parameter-parameter yang tidak diketahui, serta $K_1, K_2, \ldots, K_r$ merupakan titik knot.

2.3 Regresi Birespon Semiparametrik Spline Truncated

Analisis regresi birespon merupakan suatu analisis yang digunakan untuk mengetahui hubungan fungsional antara dua variabel respon dan variabel prediktor. Secara umum model regresi birespon dapat ditulis dalam bentuk (Wang, Guo, dan Brown, 2000):

$$\boldsymbol{y}_{ij} = \boldsymbol{h}(x_{ij}) + \boldsymbol{\varepsilon}_{ij} \qquad j=1,2 \;\; ; i = 1,2,\ldots,n \qquad (2.7)$$

dengan $\boldsymbol{y}_i = (y_{i1}, y_{i2})^T$ merupakan vektor untuk variabel respon dan $\boldsymbol{h}(x_i) = (h_1(x_i), h_2(x_i))^T$ adalah vektor dari fungsi regresi, serta $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2})^T$ adalah vektor dari error.

Regresi semiparametrik birespon adalah model regresi yang bertujuan untuk mengetahui hubungan fungsional antara dua variabel respon dengan dua atau lebih variabel prediktor, dimana hubungan antara salah satu variabel respon dengan variabel prediktor yang digunakan mengikuti pola tertentu atau disebut komponen parametrik dan hubungan antara salah satu variabel respon dengan variabel prediktor yang digunakan tidak mengikuti pola tertentu atau disebut komponen nonparametrik. Asumsi dalam regresi semiparametrik birespon yaitu terdapat korelasi antara variabel respon yang digunakan, tetapi bukan hubungan sebab akibat.

Diberikan data berpasangan $\left(x_{1i}, x_{2i}, \ldots, x_{Di}, t_{1i}, t_{2i}, \ldots, t_{Ri}, y_{ij}\right)$, $j=1,2$; $i=1,2,\cdots,n$ dengan demikian bentuk model regresi semiparametrik secara umum dapat ditulis sebagai berikut.

$$y_{ij} = f\left(x_{1i}, x_{2i}, \ldots, x_{Di}\right) + g\left(t_{1i}, t_{2i}, \ldots, t_{Ri}\right) + \varepsilon_{ij} \tag{2.8}$$

dimana y_{ij} merupakan variabel respon pertama dan kedua yang saling berkorelasi. Hubungan antara $\left(x_{1i}, x_{2i}, \ldots, x_{Di}\right)$ dan $\left(y_{i1}, y_{i2}\right)$ diasumsikan mengikuti model parametrik dengan indeks D menunjukkan banyaknya variabel prediktor komponen parametrik, sedangkan $\left(t_{1i}, \ldots, t_{Ri}\right)$ dan $\left(y_{i1}, y_{i2}\right)$ diasumsikan mengikuti model nonparametrik dengan R menunjukkan banyaknya variabel prediktor komponen nonparametrik.

Persamaan (2.8) diasumsikan bersifat aditif, sehingga diperoleh persamaan sebagai berikut.

$$\begin{aligned} y_{ij} &= f\left(x_{1j}\right) + f\left(x_{2j}\right) + \ldots + f\left(x_{Dj}\right) + g\left(t_{1j}\right) + g\left(t_{2j}\right) + \ldots + g\left(t_{Rj}\right) + \varepsilon_{ij} \\ &= \sum_{d=1}^{D} f\left(x_{di}\right) + \sum_{r=1}^{R} g\left(t_{ri}\right) + \varepsilon_{ij} \end{aligned} \tag{2.9}$$

dengan menghampiri kurva regresi $f\left(x_{di}\right)$ komponen parametrik menggunakan fungsi linier dan kurva regresi $g\left(t_{ri}\right)$ komponen nonparametrik menggunakan fungsi spline *truncated* linier dengan titik knot K sebanyak P, diperoleh penjabaran sebagai berikut.

$$y_{ij} = \beta_0 + \sum_{d=1}^{D} \beta_d x_{di} + \sum_{r=1}^{R}\left(\alpha_{r1j} t_{ri} + \sum_{p=1}^{P} \alpha_{r(1+p)\mathrm{j}}\left(t_{ri} - K_{rpj}\right)_+\right) + \varepsilon_{ij} \tag{2.10}$$

dimana $\sum_{d=1}^{D} \beta_d x_{di}$ merupakan komponen parametrik, $\sum_{r=1}^{R} \alpha_{r1j} t_{ri}$ komponen nonparametrik, dan $\sum_{p=1}^{P} \alpha_{r(1+p)\mathrm{j}}\left(t_{ri} - K_{rpj}\right)_+$ merupakan komponen *truncated*. Persamaan (2.10) dapat ditulis ke dalam bentuk matriks sebagai berikut.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{T\alpha} + \boldsymbol{\varepsilon} \tag{2.11}$$

dengan komponen-komponen pada persamaan (2.11) dijabarkan sebagai berikut.

$$\boldsymbol{y} = \begin{pmatrix} y_{11} & y_{21} & \cdots & y_{n1} & y_{12} & y_{22} & \cdots & y_{n2} \end{pmatrix}^T_{nJ \times 1}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{11} & \cdots & \beta_{D1} & \beta_{02} & \beta_{12} & \cdots & \beta_{D2} \end{pmatrix}^T_{(D+1)J\times 1}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} & \cdots & \varepsilon_{n1} & \varepsilon_{12} & \varepsilon_{22} & \cdots & \varepsilon_{n2} \end{bmatrix}^T_{nJ\times 1}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} \cdots & x_{D1} \\ 1 & x_{12} \cdots & x_{D2} \\ \vdots & \vdots \ddots & \vdots \\ 1 & x_{1n} \cdots & x_{Dn} \end{bmatrix}_{nJ\times(D+1)J}$$

$$\boldsymbol{T} = \begin{bmatrix} t_{11} & (t_{11}-K_{11j})_+ & \cdots & (t_{11}-K_{1Pj})_+ & \cdots & t_{R1} & (t_{R1}-K_{R1J})_+ & \cdots & (t_{R1}-K_{RPJ})_+ \\ t_{12} & (t_{12}-K_{11j})_+ & \cdots & (t_{12}-K_{1Pj})_+ & \cdots & t_{R2} & (t_{R2}-K_{R1J})_+ & \cdots & (t_{R2}-K_{RPJ})_+ \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ t_{1n} & (t_{1n}-K_{11j})_+ & \cdots & (t_{1n}-K_{1Pj})_+ & \cdots & t_{Rn} & (t_{Rn}-K_{R1J})_+ & \cdots & (t_{Rn}-K_{RPJ})_+ \end{bmatrix}_{nJ\times R(1+P)J}$$

2.4 Weighted Least Square (WLS)

Analisis regresi semiparametrik memiliki dua asumsi yang harus terpenuhi yaitu varian random error pada model di asumsikan homogen dan matriks varian kovarian error diketahui (Budiantara, 2010). Metode *Weighted Least Square* (WLS) digunakan apabila error tidak identik dan menyebabkan nilai varian error untuk setiap observasi tidak sama atau disebut keadaan heteroskedastisitas. Estimasi parameter menggunakan metode *Weighted Least Square* (WLS) dilakukan dengan cara meminimumkan jumlah kuadrat *error* antara pengamatan dan model yang disebut jumlah kuadrat error, dengan rumus *Metode Weighted Least Square* (WLS) adalah sebagai berikut:

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \left[R(\boldsymbol{\beta},\boldsymbol{\alpha}) \right] = \min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \left(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{T\alpha} \right)^T \boldsymbol{W} \left(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{T\alpha} \right) \tag{2.14}$$

2.5 Pemilihan Titik Knot

Titik knot merupakan titik perpaduan dimana terjadi perubahan pola perilaku fungsi. Oleh karena itu letak dan banyaknya titik knot merupakan hal penting dalam pemodelan regresi nonparametrik dengan pendekatan *spline truncated*. Wahba (1990) menyatakan bahwa salah satu metode yang digunakan untuk menentukan knot yang optimal adalah dengan metode *Generalized Cross Validation* (GCV). Metode GCV merupakan pengembangan dari metode *Cross Validation* (CV) dimana perbedaannya terletak pada faktor-faktor yang membagi residual. Pada metode GCV faktor merupakan nilai rata-rata dari faktor-faktor tersebut. Nilai GCV kemudian diperoleh dengan menjumlahkan residual-residual kuadrat yang telah terkoreksi dengan kuadrat dari faktor-faktor ini (Fitriyani, 2014). Titik knot optimal diperoleh melalui pemilihan nilai GCV terkecil (Budiantara, 2005). Fungsi GCV untuk model regresi nonparametrik multirespon *Spline Truncated* diberikan sebagai berikut:

$$GCV(k)=\frac{n^{-1}\sum_{i=1}^{n}\left(y_i-\eta\left(t_i\right)\right)^2}{\left[n^{-1}tr\left(\boldsymbol{I}\text{ - }\boldsymbol{A}\left(\boldsymbol{K}\right)\right)\right]^2} \tag{2.15}$$

dengan $\boldsymbol{A}\left(\boldsymbol{K}\right)$ merupakan suatu matriks yang memuat knot k.

2.6 Faktor-faktor yang Mempengaruhi Angka Harapan Hidup dan Pengeluaran Perkapita

Organisasi Country Health Ranking dan Roadmaps (2016) menyebutkan bahwa terdapat faktor sosial ekonomi yang mempengaruhi tingkat kesehatan, yaitu: pendidikan, pekerjaan, pendapatan, dukungan sosial dan keluarga, keamanan komunitas.

WHO menyatakan bahwa terdapat beberapa faktor (determinan) yang menentukan tingkat kesehatan, yaitu: status sosial, pendidikan, lingkungan, jaringan pendukung sosial, keturunan, dan pelayanan kesehatan. Semakin tinggi status sosial akan semakin baik tingkat kesehatan. Semakin tinggi tingkat pendidikan akan berpengaruh terhadap tingkat kesehatan yang semakain baik.

Faktor pendidikan berpengaruh terhadap tingkat kesehatan dan pendapatan seseorang. Pendidikan meningkatkan pengetahuan, kreativitas, dan imajinasi. Pendidikan juga memperluas pilihan-pilihan lain. Manusia yang berpendidikan akan lebih memperhatikan tingkat kesehatan agar dapat hidup lebih lama. Manusia yang berpendidikan akan berpeluang mendapatkan pekerjaan dan pendapatan yang layak (BPS, 2015).

Pendidikan merupakan salah satu faktor yang penting dalam pengembangan sumber daya manusia. Pendidikan bukan hanya menambah pengetahuan, tetapi juga meningkatkan keterampilan bekerja, dengan demikian dapat meningkatkan produktifitas kerja. Pendidikan dipandang sebagai investasi yang imbalannya dapat diperoleh beberapa tahun kemudian dalam bentuk pertambahan hasil kerja atau penghasilan (Simanjuntak, 2000).

Tingkat pendapatan merupakan unsur yang menentukan kemakmuran suatu masyarakat. Pendapatan masyarakat mencapai maksimum apabila kondisi tingkat penggunaan tenaga kerja penuh (*full employment*) dapat terwujud. Pengangguran akan menimbulkan efek mengurangi pendapatan masyarakat, dan itu akan mengurangi tingkat kemakmuran yang telah tercapai. Semakin turunnya tingkat kemakmuran akan menimbulkan masalah lain yaitu kemiskinan (Sukirno, 2000). Analisis faktor-faktor yang mempengaruhi daya beli (pengeluaran per kapita) masyarakat di Jawa Barat dengan menggunakan analisis regresi. Faktor-faktor yang mempengaruhi daya beli (pengeluaran per kapita) secara signifikan adalah tingkat pengangguran, PDRB sektor industri, upah minimum regional, dan tingkat inflasi Fitriani (2015).

Maully (2014) meneliti faktor-faktor yang mempengaruhi indeks kesehatan kabupaten/kota di Jawa Timur dengan menggunakan regresi logistik. Faktor yang berpengaruh signifikan terhadap indeks kesehatan di kabupaten dan kota provinsi Jawa Timur adalah persentase pertolongan pertama kelahiran pada ibu dan persentase bayi diberi imunisasi. Pemodelan angka harapan hidup di Kabupaten Jember pernah dilakukan oleh Ardianti, Wibisono, dan Jumiati (2015) dengan menggunakan metode analisis regresi berganda. Faktor yang yang berpengaruh signifikan terhadap angka harapan hidup di Kabupaten Jember adalah pendidikan, pelayanan kesehatan, perilaku hidup bersihdan sehat, serta produk domestik regional bruto (PDRB).

(Halaman ini sengaja dikosongkan)

BAB 3

METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistika provinsi Jawa Timur pada tahun 2018. Data yang digunakan pada penelitian ini meliputi unit observasi 38 kabupaten/kota di provinsi Jawa Timur.

3.2 Variabel Penelitian

Berdasarkan hasil penelitian-penelitian pada bab sebelumnya, maka variabel yang digunakan dalam penelitan ini yaitu dua variabel respon (y) dan tiga variabel prediktor (x) seperti yang disajikan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan
y_1	Angka Harapan Hidup (AHH)
y_2	Pengeluaran Perkapita
x_1	Rata-rata Lama Sekolah
x_2	Persentase Tingkat Pengangguran Terbuka (TPT)

Berikut merupakan definisi operasional dari variabel-variabel yang digunakan dalam penelitian ini.

a. Variabel respon yang digunakan dalam penelitian ini adalah angka harapan hidup dan pengeluaran perkapita tahun 2018. Angka harapan hidup adalah rata-rata tahun hidup yang akan dijalani oleh bayi yang baru lahir pada suatu tahun tertentu. Pengeluaran perkapita adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga.

b. Variabel prediktor pertama yaitu rata-rata lama sekolah. Rata-rata lama sekolah adalah jumlah tahun belajar penduduk usia 15 tahun ke atas yang telah diselesaikan dalam pendidikan formal (tidak termasuk tahun yang mengulang), menunjukkan jenjang pendidikan yang pernah atau sedang diduduki oleh seseorang.

c. Variabel prediktor kedua yaitu persentase tingkat pengangguran terbuka. Tingkat pengangguran terbuka adalah persentase jumlah angkatan kerja.yang terdiri dari mereka yang tidak punya pekerjaan dan mencari pekerjaan, mereka yang tidak punya pekerjaan dan mempersiapkan usaha, mereka yang tidak punya pekerjaan dan tidak mencari pekerjaan karena merasa tidak mungkin mendapat pekerjaan, serta mereka yang sudah punya pekerjaan tetapi belum mulai bekerja. Penduduk yang termasuk angkatan kerja adalah penduduk dengan usia kerja (15 tahun ke atas) yang bekerja, atau memiliki pekerjaan tetapi sementara tidak bekerja dan pengangguran.

Unit observasi yang digunakan dalam penelitian ini sebanyak 38 kabupaten/kota di Provinsi Jawa Timur dan banyaknya variabel prediktor terdiri dari tiga variabel. Sehingga, struktur data penelitian yang digunakan adalah sebagai berikut.

Tabel 3.2 Struktur Data Penelitian

Kabupaten/kota	**Variabel Respon**		**Variabel Prediktor**	
	y_1	y_2	x_1	x_2
1	y_{11}	y_{21}	x_{11}	x_{21}
2	y_{12}	y_{22}	x_{12}	x_{22}
3	y_{13}	y_{23}	x_{13}	x_{23}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
38	$y_{1(38)}$	$y_{2(38)}$	$x_{1(38)}$	$x_{2(38)}$

3.3 Langkah-Langkah Penelitian

Langkah-langkah untuk menjawab tujuan penelitian dalam penelitian ini yaitu sebagai berikut.

A. **Mendapatkan estimator kurva regresi semiparametrik birespon menggunakan pendekatan Spline Truncated.**

(1). Diberikan data $\left(x_{1i}, x_{2i}, \ldots, x_{Di}, t_{1i}, t_{2i}, \ldots, t_{Ri}, y_{i1}, y_{i2},\right)$, yang mengikuti model regresi semiparametrik birespon sebagai berikut:

$$y_{ij} = \sum_{d=1}^{D} c_{di}\left(x_{di}\right) + \sum_{r=1}^{R} s_{ri}(t_{ri}) + \varepsilon_{ij} \qquad j = 1,2 \; ; \; i = 1,2,\cdots,n$$

Dimana *d* adalah banyak prediktor komponen parametrik yang dihampiri dengan fungsi linier, *r* adalah banyak prediktor komponen nonparametrik yang dihampiri dengan fungsi Spline *Truncated*.

(2). Menghampiri kurva regresi komponen parametrik $c_{di}(x_{di})$ dimana $j = 1,2$; $d = 1,2,\cdots,D$ banyaknya prediktor komponen parametrik, fungsi kuadratik:

$$c_{di}(x_{di}) = \theta_0 + \sum_{d=1}^{D} \theta_{dj} x_{di} + \theta_{dj} x_{di}^2$$

(3). Menghampiri kurva regresi komponen nonparametrik $s_{ri}(t_{ri})$ dimana $r = 1,2,...,r$ banyak prediktor komponen nonparametrik menggunakan Spline Truncated linier dengan knot sebanyak *p*:

$$s_{ri}\left(t_{ri}\right) = \varphi_{0j} + \sum_{r=1}^{R} \left(\varphi_{r1j} t_{ri} + \sum_{p=1}^{P} \varphi_{r(1+p)j} \left(t_{ri} - K_{rpj}\right)_+^p \right)$$

(4). Menuliskan model regresi semiparametrik birespon linier dan kuadratik menggunakan estimator Spline Truncated dalam bentuk matriks

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{T\alpha} + \boldsymbol{\varepsilon}$$

(5). Mendapatkan estimasi untuk parameter melalui optimasi Weighted Least Square (WLS).

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \left[R\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right) \right] = \min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \left(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{T\alpha} \right)^T \boldsymbol{W} \left(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{T\alpha} \right)$$

(6). Menyelesaikan optimasi WLS menggunakan derivatif parsial.

$$\frac{\partial \min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \left[R\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right) \right]}{\boldsymbol{\beta},\boldsymbol{\alpha}} = 0$$

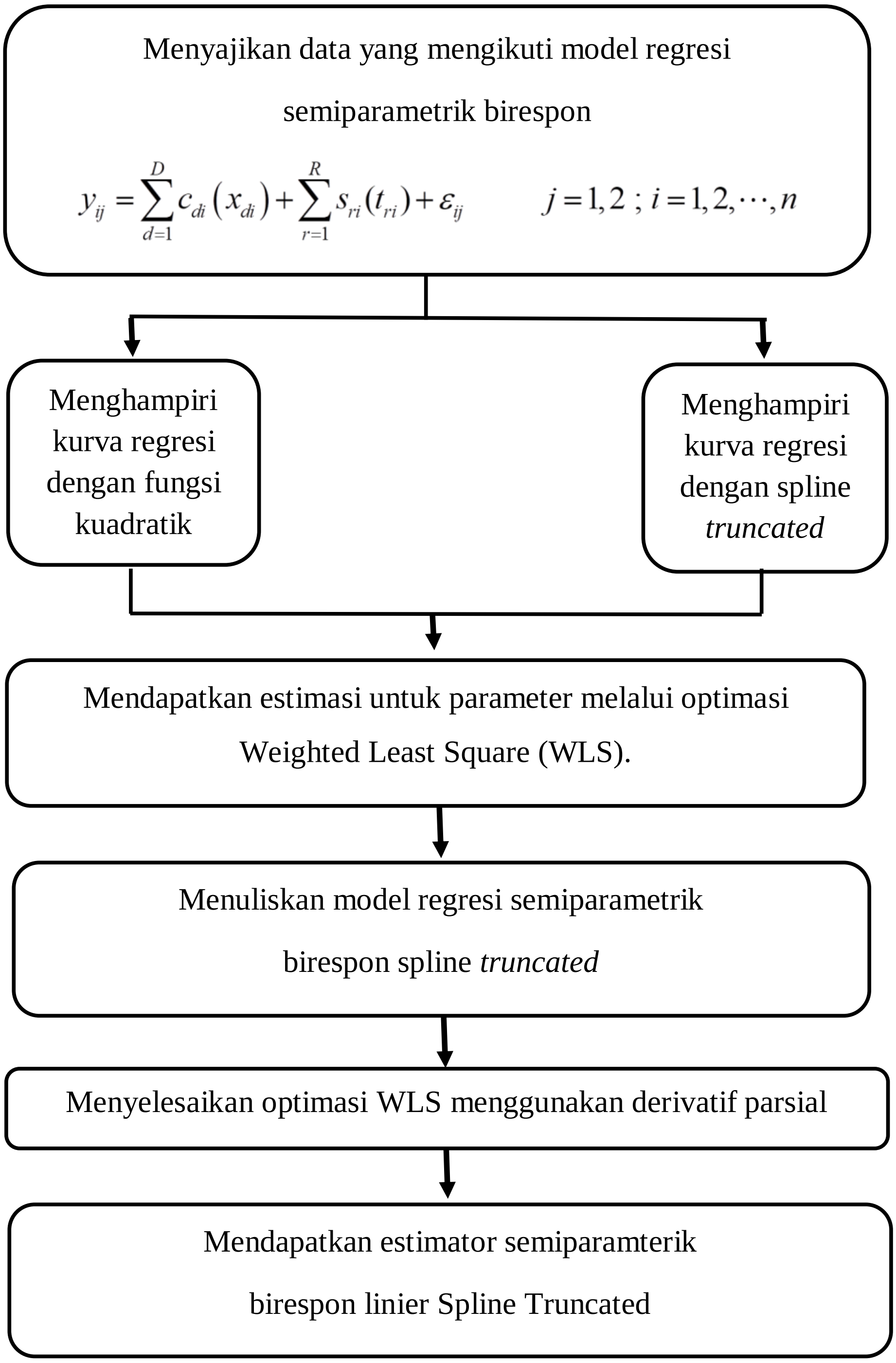
(7). Mendapatkan estimator semiparamaterik birespon linier Spline Truncated.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{T}\hat{\boldsymbol{\alpha}}$$

(8). Menulis model pada langkah (7) dalam bentuk,

$$\hat{\boldsymbol{y}} = \boldsymbol{A(K)y}$$

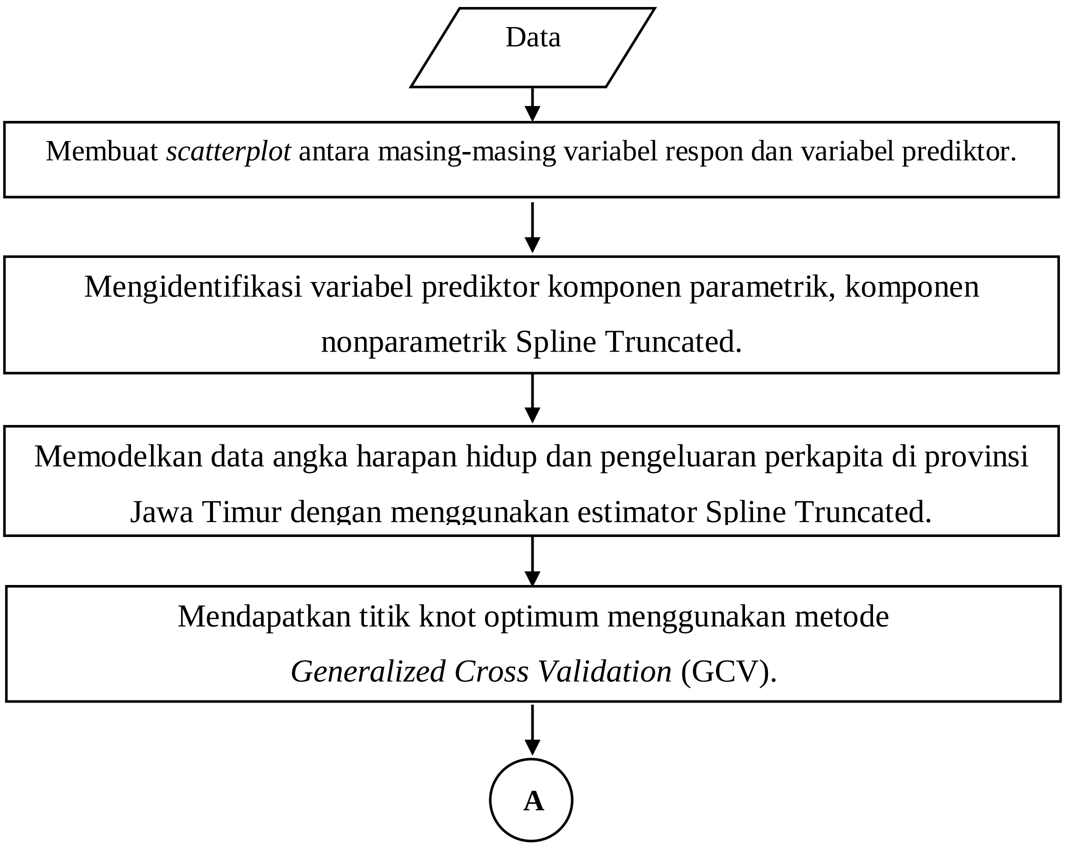
untuk suatu matriks $\boldsymbol{A(K)}$.

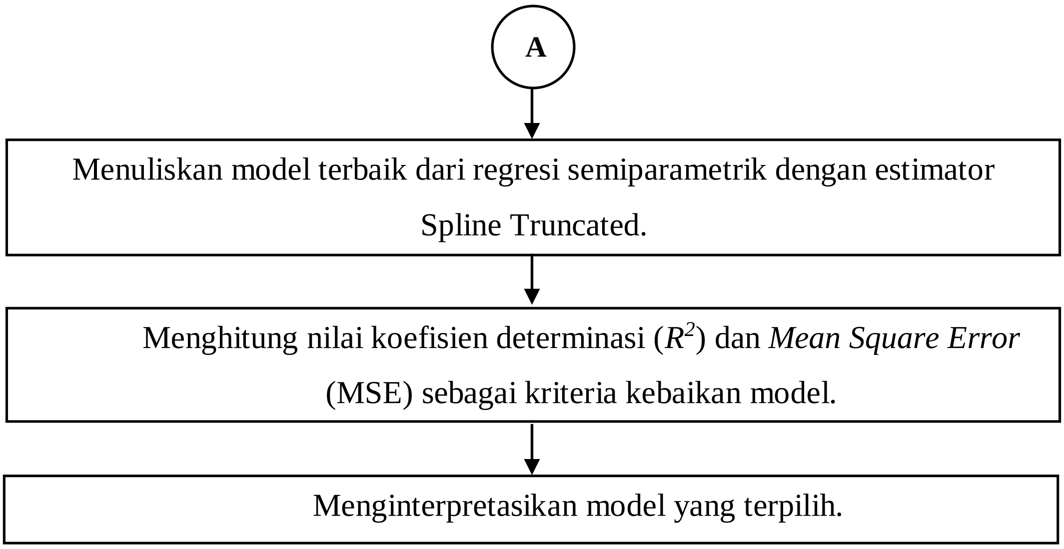


Gambar 3.1 Flow chart Estimasi Kurva Model Regresi Semiparametrik Birespon Linier dan Kuadratik menggunakan Spline *truncated*

B. Memodelkan data angka harapan hidup dan pengeluaran di bidang kesehatan di Jawa Timur menggunakan pendekatan Spline Truncated.

1. Membuat *scatterplot* antara masing-masing variabel respon dan variabel prediktor.
2. Mengidentifikasi variabel prediktor komponen parametrik, komponen nonparametrik Spline Truncated.
3. Memodelkan data angka harapan hidup dan pengeluaran perkapita di provinsi Jawa Timur dengan menggunakan estimator Spline Truncated.
4. Mendapatkan titik knot optimum menggunakan metode *Generalized Cross Validation* (GCV).
5. Menuliskan model terbaik dari regresi semiparametrik dengan estimator Spline Truncated.
6. Menghitung nilai koefisien determinasi (R^2) dan *Mean Square Error* (MSE) sebagai kriteria kebaikan model.
7. Menginterpretasikan model yang terpilih.





Gambar 3.2 *Flow chart* Memodelkan data Angka Harapan Hidup dan Pengeluaran Perkapita

BAB 4
HASIL DAN PEMBAHASAN

Pada bab ini diuraikan mengenai langkah-langkah dalam mengestimasi parameter dan kurva regresi semiparametrik birespon menggunakan pendekatan *spline truncated* dan mengaplikasikan regresi semiparametrik birespon pendekatan *spline truncated* pada data umur harapan hidup dan pengeluaran perkapita. Pada penelitian ini, estimasi parameter dilakukan terhadap dua model regresi semiparametrik birespon, yaitu regresi semiparametrik birespon yang memiliki komponen parametrik linier dan regresi semiparametrik birespon yang memiliki komponen parametrik kuadratik. Estimasi parameter pada penelitian ini diperoleh dengan menggunakan metode optimasi *Weighted Least Square* (WLS). Sementara itu untuk pemilihan titik knot optimal digunakan metode *Generalized Cross Validation* (GCV).

4.1 Estimasi Model Regresi Semiparametrik Birespon metode *Spline Truncated* dengan Komponen Parametrik Kuadratik

Diberikan data berpasangan $\left(x_{1i},\cdots,x_{Di},t_{1i},\cdots,t_{Ri},y_{ij}\right)$, j=1,2,….J ; i=1,2,$\cdots$,n hubungan pola variabel respon y_{ij} dengan variabel prediktor $x_{1i},\cdots,x_{Di}$ mengikuti pola tertentu, sehingga variabel prediktor ini disebut sebagai komponen parametrik dan hubungan variabel respon y_{ij} dengan variabel prediktor $t_{1i},t_{2i},\cdots,t_{Ri}$ tidak mengikuti pola tertentu, sehingga variabel prediktor tersebut dikatakan sebagai komponen nonparametrik dalam model regresi semiparametrik. Berdasarkan variabel-variabel yang telah dijelaskan, maka bentuk umum untuk model regresi semiparametrik birespon dengan pendekatan *spline truncated* sebagai berikut.

$$\begin{aligned} y_{ij} &= \mu\left(x_{1i},x_{2i},\cdots,x_{Di},t_{1i},t_{2i},\cdots,t_{Ri}\right)+\varepsilon_{ij} \qquad ;j=1,2\ \ ;i=1,2,\cdots n \\ &= f\left(x_{di}\right)+g\left(t_{ri}\right)+\varepsilon_{ij} \\ &= \sum_{d=1}^{D} f_{di}\left(x_{di}\right)+\sum_{r=1}^{R} g_{ri}(t_{ri})+\varepsilon_{ij} \qquad (4.1) \end{aligned}$$

dengan D adalah jumlah variabel komponen parametrik yang dihampiri dengan fungsi linier atau kuadratik, R adalah jumlah variabel prediktor yang dihampiri dengan fungsi *spline truncated*.

Estimasi parameter regresi semiparametrik birespon menggunakan metode *spline truncated* komponen parametrik linier mengacu pada penelitian yang telah dilakukan oleh Azizah (2018) dan Hidayati (2019) dengan menyesuaikan notasi yang digunakan pada penelitian ini, sehingga untuk estimasi model regresi semiparametrik birespon dengan menggunakan *spline truncated* komponen parametrik linier adalah sebagai berikut.

$$y_{ij} = \beta_0 + \sum_{d=1}^{D} \beta_d x_{di} + \sum_{r=1}^{R} \left(\alpha_{r1j} t_{ri} + \sum_{p=1}^{P} \alpha_{r(1+p)\mathrm{j}} \left(t_{ri} - K_{rpj} \right)_+ \right) + \varepsilon_{ij} \tag{4.2}$$

Selanjutnya, untuk estimasi parameter model regresi semiparametrik birespon *spline truncated* komponen kuadratik pada penelitian ini dilakukan dengan menggunakan satu variabel prediktor pada komponen parametrik. Berdasarkan persamaan (4.1) model regresi semiparametrik birespon yang memuat variabel tersebut dapat ditulis sebagai berikut.

$$\begin{aligned} y_{ij} &= \sum_{d=1}^{D} c_{di}\left(x_{di} \right) + \sum_{r=1}^{R} s_{ri}(t_{ri}) + \varepsilon_{ij} \\ &= \theta_0 + \sum_{d=1}^{D} \theta_{dj} x_{di} + \theta_{dj} x_{di}^2 + \sum_{r=1}^{R} \left(\varphi_{r1j} t_{ri} + \sum_{p=1}^{P} \varphi_{r(1+p)\mathrm{j}} \left(t_{ri} - K_{rpj} \right)_+ \right) + \varepsilon_{ij} \\ &= \theta_{0j} + \theta_{1j} x_{1i} + \theta_{2j} x_{1i}^2 + \ldots + \theta_{Dj} x_{Di} + \theta_{(D+1)j} x_{Di}^2 \\ &\quad + \varphi_{11j} t_{1i} + \varphi_{12j} \left(t_{1i} - K_{11j} \right)_+^1 + \cdots + \varphi_{1Pj} \left(t_{1i} - K_{1Pj} \right)_+^1 \\ &\quad + \cdots + \varphi_{R1j} t_{Ri} + \varphi_{R2j} \left(t_{Ri} - K_{R1j} \right)_+^1 + \cdots + \varphi_{R1j} \left(t_{Ri} - K_{RPj} \right)_+^1 + \varepsilon_{ij} \end{aligned} \tag{4.3}$$

dengan $\sum_{d=1}^{D} c_{di}\left(x_{di} \right)$ merupakan fungsi komponen parametrik yang dihampiri dengan fungsi kuadratik $\sum_{r=1}^{R} s_{ri}(t_{ri})$ merupakan fungsi spline linier dengan R knot. Persamaan (4.3) memiliki variabel respon sebanyak $j = 1, 2$ dengan banyak pengamatan $i = 1, 2, \cdots, n$ sehingga persamaan (4.3) dapat diuraikan sebagai berikut.

Untuk $j = 1$

$$\begin{aligned}
y_{11} = {} & \theta_{01} + \theta_{11}x_{11} + \theta_{21}x_{11}^2 + \ldots + \theta_{D1}x_{D1} + \theta_{D1}x_{D1}^2 + \varphi_{111}t_{11} + \varphi_{121}\left(t_{11} - K_{111}\right)_+ + \cdots + \\
& \varphi_{1P1}\left(t_{11} - K_{1P1}\right)_+ + \cdots + \varphi_{R11}t_{R1} + \varphi_{R21}\left(t_{R1} - K_{R11}\right)_+ + \cdots + \varphi_{R11}\left(t_{R1} - K_{RP1}\right)_+ + \varepsilon_{11} \\
y_{21} = {} & \theta_{01} + \theta_{11}x_{12} + \theta_{21}x_{12}^2 + \ldots + \theta_{D1}x_{D2} + \theta_{D1}x_{D2}^2 + \varphi_{111}t_{12} + \varphi_{121}\left(t_{12} - K_{111}\right)_+ + \cdots + \\
& \varphi_{1P1}\left(t_{12} - K_{1P1}\right)_+ + \cdots + \varphi_{R11}t_{R2} + \varphi_{R21}\left(t_{R2} - K_{R11}\right)_+ + \cdots + \varphi_{R11}\left(t_{R2} - K_{RP1}\right)_+ + \varepsilon_{21} \\
\vdots & \\
y_{n1} = {} & \theta_{01} + \theta_{11}x_{1n} + \theta_{21}x_{1n}^2 + \ldots + \theta_{D1}x_{Dn} + \theta_{D1}x_{Dn}^2 + \varphi_{111}t_{1n} + \varphi_{121}\left(t_{1n} - K_{111}\right)_+ + \cdots + \\
& \varphi_{1P1}\left(t_{1n} - K_{1P1}\right)_+ + \cdots + \varphi_{R11}t_{Rn} + \varphi_{R21}\left(t_{Rn} - K_{R11}\right)_+ + \cdots + \varphi_{R11}\left(t_{Rn} - K_{RP1}\right)_+ + \varepsilon_{n1}
\end{aligned}$$

Untuk $j = 2$

$$\begin{aligned}
y_{12} = {} & \theta_{02} + \theta_{12}x_{11} + \theta_{22}x_{11}^2 + \ldots + \theta_{D2}x_{D1} + \theta_{D1}x_{D1}^2 + \varphi_{112}t_{11} + \varphi_{122}\left(t_{11} - K_{112}\right)_+ + \cdots + \\
& \varphi_{1P2}\left(t_{11} - K_{1P2}\right)_+ + \cdots + \varphi_{R12}t_{R1} + \varphi_{R22}\left(t_{R1} - K_{R12}\right)_+ + \cdots + \varphi_{R12}\left(t_{R1} - K_{RP2}\right)_+ + \varepsilon_{12} \\
y_{22} = {} & \theta_{02} + \theta_{12}x_{12} + \theta_{22}x_{12}^2 + \ldots + \theta_{D2}x_{D2} + \theta_{D2}x_{D2}^2 + \varphi_{112}t_{12} + \varphi_{122}\left(t_{12} - K_{112}\right)_+ + \cdots + \\
& \varphi_{1P2}\left(t_{12} - K_{1P2}\right)_+ + \cdots + \varphi_{R12}t_{R2} + \varphi_{R22}\left(t_{R2} - K_{R12}\right)_+ + \cdots + \varphi_{R12}\left(t_{R2} - K_{RP2}\right)_+ + \varepsilon_{22} \\
\vdots & \\
y_{n2} = {} & \theta_{02} + \theta_{12}x_{1n} + \theta_{22}x_{1n}^2 + \ldots + \theta_{D2}x_{Dn} + \theta_{D1}x_{Dn}^2 + \varphi_{112}t_{1n} + \varphi_{122}\left(t_{1n} - K_{112}\right)_+ + \cdots + \\
& \varphi_{1P2}\left(t_{1n} - K_{1P2}\right)_+ + \cdots + \varphi_{R12}t_{Rn} + \varphi_{R22}\left(t_{Rn} - K_{R12}\right)_+ + \cdots + \varphi_{R12}\left(t_{Rn} - K_{RP2}\right)_+ + \varepsilon_{n2}
\end{aligned}$$

Model persamaan (4.3) setelah diuraikan, selanjutnya dapat dibentuk menjadi sebuah matriks sebagaimana berikut.

$$\begin{bmatrix} \boldsymbol{y_1} \\ --- \\ \boldsymbol{y_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_1\theta_1} \\ --- \\ \boldsymbol{X_2\theta_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{T_1\varphi_1} \\ -- \\ \boldsymbol{T_2\varphi_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ -- \\ \boldsymbol{\varepsilon_2} \end{bmatrix} \tag{4.4}$$

Persamaan (4.4) dapat ditulis kembali dalam bentuk matriks:

$$\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{T\varphi} + \boldsymbol{\varepsilon} \tag{4.5}$$

Persamaan (4.5) tersusun dari matriks dan vektor untuk $i = 1, 2, \cdots, n$ dengan penjabaran sebagai berikut.

Vektor-vektor penyusun vektor $\boldsymbol{y}$

$$\boldsymbol{y}_1 = \begin{bmatrix} y_{11} & y_{21} & \ldots & y_{n1} \end{bmatrix}^T_{n\times 1}$$

$$\boldsymbol{y}_2 = \begin{bmatrix} y_{12} & y_{22} & \ldots & y_{n2} \end{bmatrix}^T_{n\times 1}$$

Vektor-vektor penyusun vektor parameter θ dan φ.

$$\boldsymbol{\theta}_1 = \left[\theta_{01} \quad \theta_{11} \quad \theta_{21} \quad \cdots \quad \theta_{D1}\right]^T_{(D+1)\times 1}$$

$$\boldsymbol{\theta}_2 = \left[\theta_{02} \quad \theta_{12} \quad \theta_{22} \quad \cdots \quad \theta_{D2}\right]^T_{(D+1)\times 1}$$

$$\boldsymbol{\varphi}_1 = \left[\varphi_{111} \quad \varphi_{121} \quad \cdots \quad \varphi_{1P1} \cdots \quad \varphi_{R11} \quad \varphi_{R21} \quad \cdots \quad \varphi_{RP1}\right]^T_{R(1+P)\times 1}$$

$$\boldsymbol{\varphi}_2 = \left[\varphi_{112} \quad \varphi_{122} \quad \cdots \quad \varphi_{1P2} \cdots \quad \varphi_{R12} \quad \varphi_{R22} \quad \cdots \quad \varphi_{RP2}\right]^T_{R(1+P)\times 1}$$

Matriks penyusun matriks prediktor $\boldsymbol{X}$ dan $\boldsymbol{T}$ adalah sebagai berikut.

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \cdots & x_{D1} & x_{D1}^2 \\ 1 & x_{12} & x_{12}^2 & \cdots & x_{D2} & x_{D2}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1n} & x_{1n}^2 & \cdots & x_{Dn} & x_{Dn}^2 \end{bmatrix}_{n\times(D+2)}$$

$$\boldsymbol{X}_2 = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \cdots & x_{D1} & x_{D1}^2 \\ 1 & x_{12} & x_{12}^2 & \cdots & x_{D2} & x_{D2}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1n} & x_{1n}^2 & \cdots & x_{Dn} & x_{Dn}^2 \end{bmatrix}_{n\times(D+2)}$$

$$\boldsymbol{T}_1 = \begin{bmatrix} t_{11} & (t_{11}-K_{111})_+ & \cdots & (t_{11}-K_{1P1})_+ & \cdots & t_{R1} & (t_{R1}-K_{R11})_+ & \cdots & (t_{R1}-K_{RP1})_+ \\ t_{12} & (t_{12}-K_{111})_+ & \cdots & (t_{12}-K_{1P1})_+ & \cdots & t_{R2} & (t_{R2}-K_{R11})_+ & \cdots & (t_{R2}-K_{RP1})_+ \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ t_{1n} & (t_{1n}-K_{111})_+ & \cdots & (t_{1n}-K_{1P1})_+ & \cdots & t_{Rn} & (t_{Rn}-K_{R11})_+ & \cdots & (t_{Rn}-K_{RP1})_+ \end{bmatrix}_{n\times R(1+P)}$$

$$\boldsymbol{T}_2 = \begin{bmatrix} t_{11} & (t_{11}-K_{112})_+ & \cdots & (t_{11}-K_{1P2})_+ & \cdots & t_{R1} & (t_{R1}-K_{R12})_+ & \cdots & (t_{R1}-K_{RP2})_+ \\ t_{12} & (t_{12}-K_{112})_+ & \cdots & (t_{12}-K_{1P2})_+ & \cdots & t_{R2} & (t_{R2}-K_{R12})_+ & \cdots & (t_{R2}-K_{RP2})_+ \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ t_{1n} & (t_{1n}-K_{112})_+ & \cdots & (t_{1n}-K_{1P2})_+ & \cdots & t_{Rn} & (t_{Rn}-K_{R12})_+ & \cdots & (t_{Rn}-K_{RP2})_+ \end{bmatrix}_{n\times R(1+P)}$$

Vektor-vektor penyusun vektor $\boldsymbol{\varepsilon}$ adalah sebagai berikut.

$$\boldsymbol{\varepsilon}_1 = \left[\varepsilon_{11} \quad \varepsilon_{21} \quad \cdots \quad \varepsilon_{n1}\right]^T_{n\times 1}$$

$$\boldsymbol{\varepsilon}_2 = \left[\varepsilon_{12} \quad \varepsilon_{22} \quad \cdots \quad \varepsilon_{n2}\right]^T_{n\times 1}$$

Secara umum, persamaan (4.5) dapat ditulis dalam bentuk berikut.

$$\boldsymbol{y}_{nJ\times 1} = \boldsymbol{X}_{nJ\times(D+2)J}\boldsymbol{\theta}_{(D+2)J\times 1} + \boldsymbol{T}_{nJ\times R(1+P)J}\boldsymbol{\varphi}_{R(1+P)J\times 1} + \boldsymbol{\varepsilon}_{nJ\times 1}$$

dimana

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 \end{bmatrix} \quad ; \quad \boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_2 \end{bmatrix}$$

dengan $\boldsymbol{0}$ merupakan matriks yang semua elemennya bernilai nol dengan ukuran $\boldsymbol{0}_{n\times(D+2)}$ pada matriks $\boldsymbol{X}$ dan $\boldsymbol{0}_{n\times R(1+P)}$ pada matriks $\boldsymbol{T}$.

Perbedaan estimasi model regresi semiparametrik birespon menggunakan *spline truncated* untuk komponen parametrik kuadratik dengan komponen parametrik linier terletak pada struktur penyusun matriks variabel prediktor. Pada komponen parametrik kuadratik, terdapat satu penambahan kolom variabel prediktor yang menyebabkan ukuran dari matriks $\boldsymbol{X}$ berubah atau berbeda dengan ukuran dari matriks $\boldsymbol{X}$ pada komponen parametrik linier.

Selanjutnya, estimasi parameter $\tilde{\theta}$ dan $\tilde{\varphi}$ dilakukan dengan metode dan langkah yang sama seperti estimasi dengan komponen parametrik linier yaitu menggunakan metode *Weighted Least Square* (WLS). Karena adanya syarat terpenuhinya korelasi antara variabel respon satu dengan dua, maka dibutuhkan matriks varian kovarian $\boldsymbol{W}$ dalam mengestimasi regresi semiparametrik birespon berdasarkan estimator *spline truncated*. Matriks $\boldsymbol{W}$ adalah matriks pembobot (varian kovarian) variabel y dengan struktur sebagai berikut.

$$\boldsymbol{W}_{2n \text{ x } 2n} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}_{2\text{x}2} \otimes \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}_{n \text{ x } n}$$

$$\sigma_{11} = \text{var}(y_1) = \frac{\sum_{i=1}^{n}(y_{1i} - \bar{y}_1)^2}{n-1}$$

$$\sigma_{22} = \text{var}(y_2) = \frac{\sum_{i=1}^{n}(y_{2i} - \bar{y}_1)^2}{n-1}$$

$$\sigma_{12} = \sigma_{21} = \text{covar}(y_1, y_2) = \sum_{i=1}^{n}(y_{1i} - \bar{y}_1)^2 (y_{2i} - \bar{y}_2)^2$$

Berdasarkan persamaan (4.5), maka diperoleh.

$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\theta} - \boldsymbol{T\varphi}$

Sehingga, dengan meminimumkan kriteria *Weighted Least Square* (WLS) maka diperoleh estimator parameter θ dan φ dengan penjabaran sebagai berikut.

$$
\begin{aligned}
L &= \boldsymbol{\varepsilon}^T \boldsymbol{W}\boldsymbol{\varepsilon} \\
&= \left(\boldsymbol{y} - \boldsymbol{X\theta} - \boldsymbol{T\varphi}\right)^T \boldsymbol{W} \left(\boldsymbol{y} - \boldsymbol{X\theta} - \boldsymbol{T\varphi}\right) \\
&= \left(\boldsymbol{y}^T - \boldsymbol{\theta}^T \boldsymbol{X}^T - \boldsymbol{\varphi}^T \boldsymbol{T}^T\right)\left(\boldsymbol{Wy} - \boldsymbol{WX\theta} - \boldsymbol{WT\varphi}\right) \\
&= \boldsymbol{y}^T \boldsymbol{Wy} - \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{Wy} - \boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{Wy} - \boldsymbol{y}^T \boldsymbol{WX\theta} + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{WX\theta} \\
&\quad + \boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WX\theta} - \boldsymbol{y}^T \boldsymbol{WT\varphi} + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{WT\varphi} + \boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WT\varphi} \\
&= \boldsymbol{y}^T \boldsymbol{Wy} - 2\boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{Wy} - 2\boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{Wy} + 2\boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WX\theta} \\
&\quad + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{WX\theta} + \boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WT\varphi}
\end{aligned}
$$

Estimator untuk θ dan φ diperoleh dengan melakukan derivatif parsial dari L, dengan penjabaran derivatif parsial terhadap θ adalah sebagai berikut:

$$
\frac{\partial L}{\partial \boldsymbol{\theta}} = \frac{\partial\left(\boldsymbol{y}^T \boldsymbol{Wy} - 2\boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{Wy} - 2\boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{Wy} + 2\boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WX\theta} + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{WX\theta} + \boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WT\varphi}\right)}{\partial \boldsymbol{\theta}}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\theta}} &= 0 \\
-2\boldsymbol{X}^T \boldsymbol{Wy} + 2\boldsymbol{X}^T \boldsymbol{WT\varphi} + 2\boldsymbol{X}^T \boldsymbol{WX\theta} &= 0 \\
-\boldsymbol{X}^T \boldsymbol{Wy} + \boldsymbol{X}^T \boldsymbol{WT\varphi} + \boldsymbol{X}^T \boldsymbol{WX\theta} &= 0 \\
\boldsymbol{X}^T \boldsymbol{WX\theta} &= \boldsymbol{X}^T \boldsymbol{Wy} - \boldsymbol{X}^T \boldsymbol{WT\varphi}
\end{aligned}
$$

sehingga, diperoleh:

$$
\hat{\boldsymbol{\theta}} = \left(\boldsymbol{X}^T \boldsymbol{WX}\right)^{-1} \left(\boldsymbol{X}^T \boldsymbol{Wy} - \boldsymbol{X}^T \boldsymbol{WT}\hat{\boldsymbol{\varphi}}\right) \tag{4.6}
$$

Selanjutnya penjabaran derivatif parsial terhadap φ adalah sebagai berikut:

$$
\frac{\partial L}{\partial \boldsymbol{\varphi}} = \frac{\partial\left(\boldsymbol{y}^T \boldsymbol{Wy} - 2\boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{Wy} - 2\boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{Wy} + 2\boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WX\theta} + \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{WX\theta} + \boldsymbol{\varphi}^T \boldsymbol{T}^T \boldsymbol{WT\varphi}\right)}{\partial \boldsymbol{\varphi}}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\varphi}} &= 0 \\
-2\boldsymbol{T}^T \boldsymbol{Wy} + 2\boldsymbol{T}^T \boldsymbol{WX\theta} + 2\boldsymbol{T}^T \boldsymbol{WT\varphi} &= 0 \\
-\boldsymbol{T}^T \boldsymbol{Wy} + \boldsymbol{T}^T \boldsymbol{WX\theta} + \boldsymbol{T}^T \boldsymbol{WT\varphi} &= 0 \\
\boldsymbol{T}^T \boldsymbol{WT\varphi} &= \boldsymbol{T}^T \boldsymbol{Wy} - \boldsymbol{T}^T \boldsymbol{WX\theta}
\end{aligned}
$$

sehingga diperoleh:

$$
\hat{\boldsymbol{\varphi}} = \left(\boldsymbol{T}^T \boldsymbol{WT}\right)^{-1} \left(\boldsymbol{T}^T \boldsymbol{Wy} - \boldsymbol{T}^T \boldsymbol{WX}\hat{\boldsymbol{\theta}}\right) \tag{4.7}
$$

Estimator parameter $\hat{\theta}$ pada persamaan (4.6) masih memuat estimator parameter $\hat{\boldsymbol{\varphi}}$. Demikian pula pada estimator parameter $\hat{\varphi}$ pada persamaan (4.7) masih memuat estimator parameter $\hat{\boldsymbol{\theta}}$, sehingga untuk memperoleh estimator parameter yang saling bebas maka perlu dilakukan subtitusi pada persamaan (4.7) dan (4.6).

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\hat{\boldsymbol{\varphi}}\right) \\
&= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\left[\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left\{\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\theta}}\right)\right\}\right] \\
&= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\left[\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\theta}}\right] \\
&= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y} + \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\theta}}
\end{aligned}$$

Selanjutnya, untuk suku yang memuat estimator $\hat{\theta}$ dikelompokkan dalam satu ruas.

$$\hat{\boldsymbol{\theta}} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\theta}} = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y}$$

$$\hat{\boldsymbol{\theta}}\left[\mathbf{I} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\right] = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y}$$

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{I} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\left(\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y}\right)$$

Misalkan, $\mathbf{M} = \left(\mathbf{I} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}$ maka, diperoleh estimator $\hat{\theta}$ untuk komponen parametrik dalam model regresi semiparametrik birespon adalah sebagai berikut.

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left(\mathbf{I} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1} \\
&\quad \left(\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y}\right) \\
&= \mathbf{M}\left(\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y}\right) \\
&= \mathbf{M}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} - \mathbf{M}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{y} \\
&= \mathbf{M}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\left\{\boldsymbol{X}^T - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\right\}\boldsymbol{W}\boldsymbol{y} \\
&= \mathbf{C}(\boldsymbol{K})\boldsymbol{y} \qquad (4.8)
\end{aligned}$$

Dengan matriks hat untuk komponen parametrik dalam model adalah:

$$\boldsymbol{C}(\boldsymbol{K}) = \mathbf{M}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\left\{\boldsymbol{X}^T - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{T}\left(\boldsymbol{T}^T\boldsymbol{W}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^T\right\}\boldsymbol{W}$$

dimana, matriks $\boldsymbol{C}(\boldsymbol{K})$ merupakan matriks yang terdiri dari variabel prediktor untuk komponen parametrik.

Selanjutnya, untuk memperoleh estimator $\hat{\varphi}$ yang bebas dari estimator $\hat{\theta}$ dilakukan cara yang sama yaitu dengan menstubtitusikan persamaan (4.6) ke dalam persamaan (4.7) sebagai berikut:

$$\begin{aligned}
\hat{\varphi} &= \left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left(\boldsymbol{T}^T\boldsymbol{Wy}-\boldsymbol{T}^T\boldsymbol{WX}\hat{\theta}\right)\\
&= \left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left[\boldsymbol{T}^T\boldsymbol{Wy}-\boldsymbol{T}^T\boldsymbol{WX}\left\{\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{Wy}-\boldsymbol{X}^T\boldsymbol{WT}\hat{\varphi}\right)\right\}\right]\\
&= \left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left[\boldsymbol{T}^T\boldsymbol{Wy}-\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}-\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\hat{\varphi}\right]\\
&= \left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}+\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\hat{\varphi}
\end{aligned}$$

Selanjutnya, untuk suku yang memuat estimator $\hat{\varphi}$ dikelompokkan dalam satu ruas:

$$\hat{\varphi}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\hat{\varphi}=\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}$$

$$\hat{\varphi}\left(\mathbf{I}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\right)=\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}$$

$$\hat{\varphi}=\left(\mathbf{I}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\right)^{-1}\left(\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}\right)$$

Misalkan, $\mathbf{N}=\left(\mathbf{I}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\right)^{-1}$ maka, didapat estimator $\hat{\varphi}$ untuk komponen nonparametrik dalam model regresi semiparametrik birespon adalah sebagai berikut:

$$\begin{aligned}
\hat{\varphi} &= \left(\mathbf{I}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{WT}\right)^{-1}\\
&\quad\left(\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}\right)\\
&= \mathbf{N}\left(\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}\right)\\
&= \mathbf{N}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{Wy}-\mathbf{N}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Wy}\\
&= \mathbf{N}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left\{\boldsymbol{T}^T-\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\right\}\boldsymbol{Wy}\\
&= \mathbf{D}\boldsymbol{(K)y}
\end{aligned} \tag{4.9}$$

Dengan matriks hat untuk komponen parametrik dalam model adalah:

$$\boldsymbol{D(K)}=\mathbf{N}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left\{\boldsymbol{T}^T-\boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\right\}\boldsymbol{W}$$

dimana, matriks $\boldsymbol{D(K)}$ merupakan matriks yang terdiri dari variabel prediktor untuk komponen nonparametrik.

Sehingga, berdasarkan persamaan (4.8) dan (4.9) diperoleh model regresi semiparametrik birespon kuadratik dengan pendekatan *spline truncated* adalah sebagai berikut:

$$\begin{aligned}\hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\boldsymbol{\theta}} + \boldsymbol{T}\hat{\boldsymbol{\varphi}} \\ &= \boldsymbol{XC(K)y} + \boldsymbol{TD(K)y} \\ &= \left(\boldsymbol{XC(K)} + \boldsymbol{TD(K)}\right)\boldsymbol{y} \\ &= \boldsymbol{A(K)y}\end{aligned} \qquad (4.10)$$

dengan

$$\boldsymbol{A(K)} = \left(\mathbf{M}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\left\{\boldsymbol{X}^T - \boldsymbol{X}^T\boldsymbol{WT}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\right\}\boldsymbol{W}\right) + \left(\mathbf{N}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left\{\boldsymbol{T}^T - \boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\right\}\boldsymbol{W}\right)$$

dimana, $\boldsymbol{A(K)}$ merupakan matriks yang terdiri dari variabel prediktor untuk komponen parametrik dan nonparametrik serta memuat titik-titik knot.

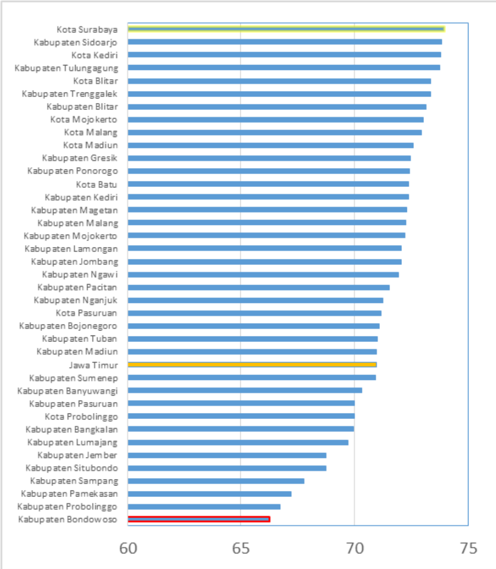
Berdasarkan ulasan yang telah diberikan, tujuan pertama dalam penelitian ini telah tercapai.

4.2 Aplikasi pada Data Angka Harapan Hidup dan Pengeluaran Perkapita di Jawa Timur

Pada bagian ini dibahas mengenai parameter model spline truncated regresi semiparametrik birespon komponen parametrik linier dan kuadratik yang diaplikasikan pada data Angka Harapan Hidup (AHH) dan Pengeluaran Perkapita di Jawa Timur dengan menggunakan *software* R (*syntax* berada pada Lampiran 2).

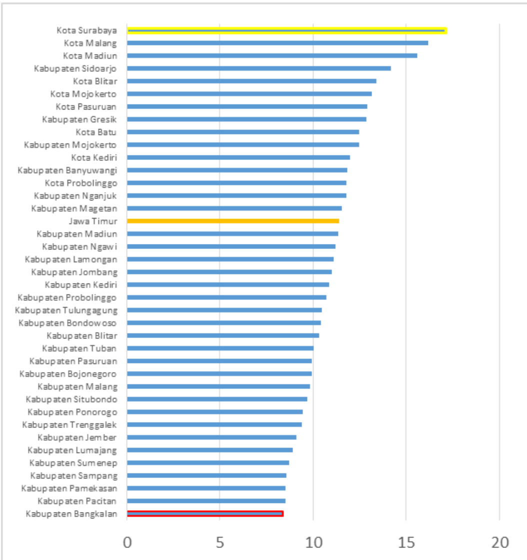
4.2.1 Karakteristik Data

Karakteristik data diperlukan untuk eksplorasi data dan sebagai gambaran umum tentang variabel yang digunakan dalam melakukan pemodelan dan estimasi parameter regresi semiparametrik birespon linier dan kuadratik menggunakan metode *spline truncated.*



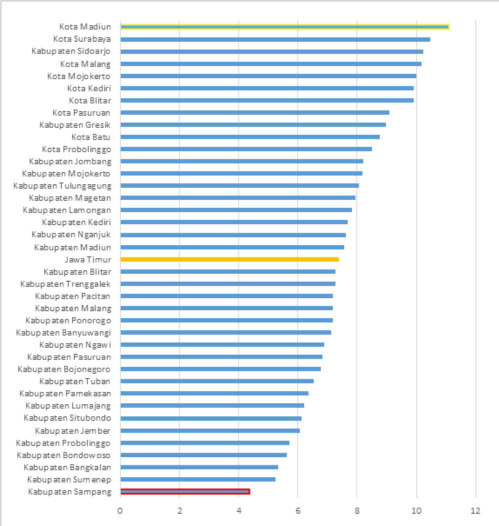
Gambar 4.1 Karakeristik Data Angka Harapan Hidup

Berdasarkan Gambar 4.1, Jawa Timur pada tahun 2018 memiliki nilai AHH sebesar 70.97 persen, nilai tersebut berada di bawah nilai AHH nasional yaitu sebesar 71.20 persen. Terdapat 17 kab/kota di Jawa Timur yang memiliki nilai AHH di bawah AHH nasional, dan 21 kab/kota yang memiliki nilai AHH di atas AHH nasional. Nilai tertinggi berada di Kota Surabaya sedangkan terendah berada di kabupaten Bondowoso.



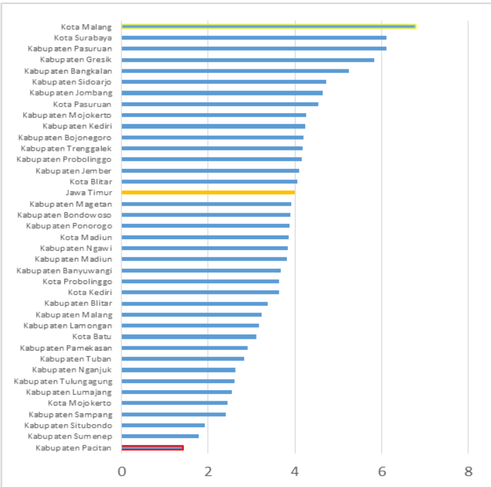
Gambar 4.2 Karakteristik Data Pengeluaran Perkapita

Gambar 4.2 menunjukkan bahwa Jawa Timur pada tahun 2018 memiliki pengeluaran perkapita sebesar 11.38, nilai tersebut berada di atas nilai pengeluaran perkapita nasional yaitu sebesar 11.05. Terdapat 20 kab/kota di Jawa Timur yang memiliki nilai pengeluaran perkapita di bawah pengeluaran perkapita nasional, dan 18 kab/kota yang memiliki nilai pengeluaran perkapita di atas pengeluaran perkapita nasional. Pengeluaran perkapita tertitnggi berada di kota Surabaya sedangkan untuk pengeluaran perkapita terendah berada di kabupaten Bangkalan.



Gambar 4.3 Karakteristik Data Rata-rata Lama Sekolah

Berdasarkan Gambar 4.3 didapatkan rata-rata lama sekolah di Jawa Timur sebesar 7.39 tahun, nilai tersebut berada di bawah nilai rata-rata lama sekolah nasional yaitu sebesar 8.17 tahun. Terdapat 25 kab/kota di Jawa Timur yang memiliki nilai rata-rata lama sekolah di bawah rata-rata lama sekolah nasional, dan 13 kab/kota yang memiliki nilai rata-rata pengeluaran perkapita di atas rata-rata lama sekolah nasional. Kota Madiun memiliki rata-rata lama sekolah tertinggi sedangkan rata-rata lama sekolah terendah berada di daerah Sampang.



Gambar 4.4 Karakteristik Data Tingkat Pengangguran Terbuka

Gambar 4.4 Jawa Timur pada tahun 2018 memiliki nilai TPT sebesar 3.99 persen, nilai tersebut berada di bawah nilai nilai TPT nasional yaitu sebesar 5.34 persen. Terdapat 34 kab/kota di Jawa Timur yang memiliki nilai nilai TPT di bawah nilai TPT nasional, dan 4 kab/kota yang memiliki nilai TPT di atas nilai TPT nasional. Nilai TPT tertinggi terdapat di kota Malang dan kabupaten Pacitan sebagai daerah yang memiliki TPT terendah. Tabel 4.1 menunjukkan nilai tertinggi dan terendah dari masing-masing variabel.

Tabel 4.1 Statistika Deskriptif Variabel Penelitian

Variabel	Minimum	Maximum
AHH (y_1)	66.27	73.98
Pengeluaran Perkapita (y_2)	8.39	17.15
Rata-rata lama sekolah (x_1)	1.43	6.79
TPT(x_2)	4.36	11.11

4.2.2 Uji Korelasi Variabel Respon

Pengujian korelasi pada variabel respon yang digunakan dalam penelitian perlu dilakukan untuk mengetahui seberapa besar hubungan antara variabel respon AHH dengan pengeluaran perkapita secara statistik. Metode uji yang digunakan adalah uji Korelasi Pearson, dengan hipotesis pengujian sebagai berikut.

$H_0 : \rho_{12} = 0$ (Tidak ada korelasi antara AHH dengan Pengeluaran perkapita)

$H_1 : \rho_{12} \neq 0$ (Terdapat korelasi antara AHH dengan Pengeluaran perkapita)

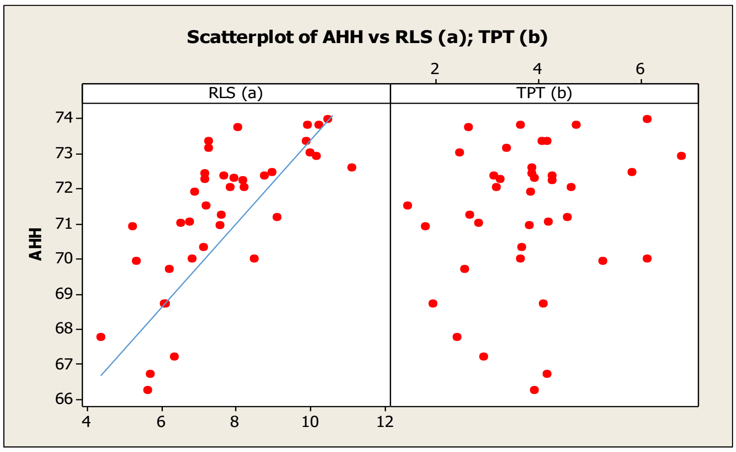
Berdasarkan pengujian korelasi, didapatkan hasil pada Lampiran 3 yang menunjukkan hasil nilai korelasi Pearson sebesar 0.527 dan nilai P-*value* sebesar 0.001 kurang dari taraf signifikansi $\alpha = 0.01$ sehingga dapat disimpulkan bahwa Angka Harapan Hidup dan pengeluaran perkapita saling berkorelasi atau independen.

4.2.3 Identifikasi Variabel Komponen Parametrik dan Nonparametrik

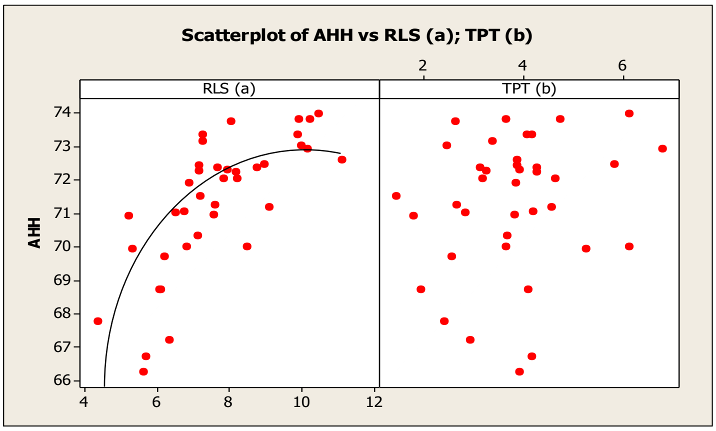
Identifikasi variabel komponen parametrik dan nonparametrik dilakukan dengan melihat pola data (*scatter plot*) antara variabel respon dengan variabel prediktor. Dalam regresi nonparametrik, *scatter plot* memiliki perubahan pola perilaku yang dimana lokasi perubahan pola tersebut disebut titik knot. Pola perubahan *scatter plot* dapat dikatakan mengikuti pola data nonparametrik *spline truncated* apabila (Budiantara, 2019):

1. Pola data yang awalnya naik, berubah menjadi pola data turun dititik knotnya.
2. Pola data yang awalnya turun, berubah menjadi pola data naik dititik knotnya.
3. Pola data yang awalnya naik dengan gradien cepat (lambat) berubah menjadi pola data tetap naik di titik knotnya dengan lambat (cepat).
4. Pola data yang diasumsikan tidak mengikuti pola tertentu.

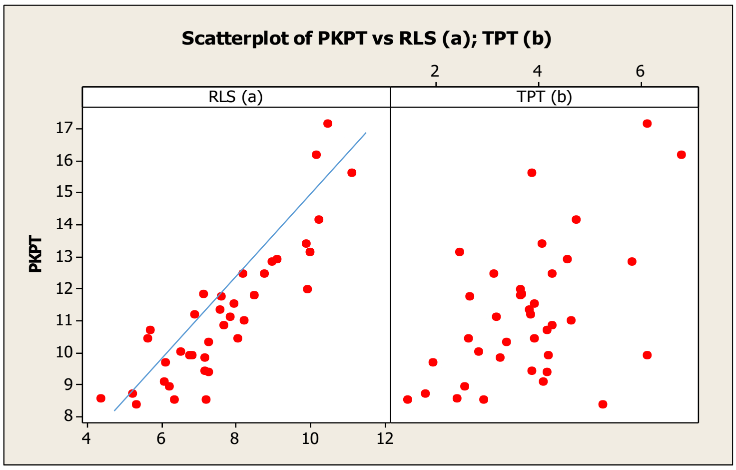
Pola hubungan data antara variabel respon dan variabel prediktor ditunjukkan oleh *scatter plot* pada Gambar 4.5, Gambar 4.6., Gambar 4.7., dan Gambar 4.8.



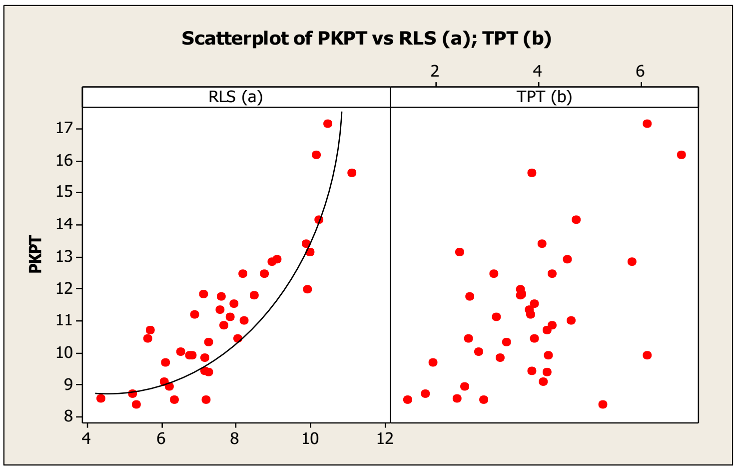
Gambar 4.5 *Scatter Plot* Angka Harapan Hidup dengan variabel prediktor Rata-rata Lama Sekolah (a) dan Tingkat Pengangguran Terbuka (b) dengan pendekatan komponen parametrik linier.



Gambar 4.6 *Scatter Plot* Angka Harapan Hidup dengan variabel prediktor Rata-rata Lama Sekolah (a) dan Tingkat Pengangguran Terbuka (b) dengan pendekatan komponen parametrik kuadratik.



Gambar 4.7 *Scatter Plot* Pengeluaran perkapita dengan variabel prediktor Rata-rata Lama Sekolah (a) dan Tingkat Pengangguran Terbuka (b) dengan pendekatan komponen parametrik linier.



Gambar 4.8 *Scatter Plot* Pengeluaran perkapita dengan variabel prediktor Rata-rata Lama Sekolah (a) dan Tingkat Pengangguran Terbuka (b) dengan pendekatan komponen parametrik kuadratik.

Berdasarkan Gambar 4.5 pola hubungan antara variabel respon AHH dengan variabel prediktor rata-rata lama sekolah cenderung membentuk pola parametrik. Gambar 4.7 juga menunjukkan pola hubungan yang cenderung membentuk pola parametrik antara variabel respon pengeluaran perkapita dengan variabel prediktor rata-rata lama sekolah. Berdasarkan Gambar 4.6 dan 4.7 pola hubungan antara variabel respon AHH dengan rata-rata lama sekolah dan variabel respon pengeluaran perkapita dengan rata-rata lama sekolah dapat diasumsikan mengikuti pola kuadratik, sehingga berdasarkan alasan tersebut, maka diasumsikan bahwa rata-rata lama sekolah sebagai variabel komponen parametrik (linier atau kuadratik) yang disimbolkan dengan x. Berdasarkan Gambar 4.5, Gambar 4.6, Gambar 4.7, dan Gambar 4.8 pola hubungan antara variabel respon AHH dengan variabel prediktor TPT dan variabel respon pengeluaran perkapita dengan variabel prediktor TPT cenderung tidak membentuk pola tertentu, sehingga variabel prediktor TPT dikatakan sebagai variabel komponen nonparametrik yang disimbolkan dengan t.

Berdasarkan uraian sebelumnya yaitu persamaan (4.2), maka model regresi semiparametrik birespon *spline truncated* dengan komponen parametrik linier dapat dituliskan sebagai berikut.

Respon 1 (Angka Harapan Hidup)

$$\hat{y}_{1i} = \hat{\beta}_{01} + \hat{\beta}_{11}x_i + \hat{\alpha}_{11}t_{1i} + \hat{\alpha}_{111}\left(t_{1i} - K_{111}\right)_+ \tag{4.11}$$

Respon 2 (Pengeluaran per kapita)

$$\hat{y}_{2i} = \hat{\beta}_{02} + \hat{\beta}_{12}x_i + \hat{\alpha}_{21}t_{1i} + \hat{\alpha}_{211}\left(t_{1i} - K_{211}\right)_+ \tag{4.12}$$

Untuk model regresi semiparametrik birespon *spline truncated* komponen parametrik kuadratik dapat dituliskan sebagai berikut.

Respon 1 (Angka Harapan Hidup)

$$\hat{y}_{1i} = \hat{\theta}_{01} + \hat{\theta}_{11}x_1 + \hat{\theta}_{21}x_i^{\,2} + \hat{\varphi}_{11}t_{1i} + \hat{\varphi}_{111}\left(t_{1i} - K_{111}\right)_+ \tag{4.13}$$

Respon 2 (Pengeluaran per kapita)

$$\hat{y}_{2i} = \hat{\theta}_{02} + \hat{\theta}_{12}x_i + \hat{\theta}_{22}x_i^{\,2} + \hat{\varphi}_{11}t_{1i} + \hat{\varphi}_{111}\left(t_{1i} - K_{211}\right)_+ \tag{4.14}$$

Dengan fungsi *spline truncated* diberikan pada persamaan:

$$\left(t_i - K_{ri}\right)_+ = \begin{cases} \left(t_i - K_{ri}\right) & , t_i \geq K_{ri} \\ 0 & , t_i < K_{ri} \end{cases}$$

Selanjutnya dilakukan pemilihan titik knot optimal untuk variabel komponen nonparametrik *spline truncated* menggunakan metode GCV minimum.

4.2.4 Pemilihan Titik Knot Optimal

Pemilihan titik knot optimal pada regresi semiparametrik birespon *spline truncated* komponen parametrik linier dan kuadratik dengan satu variabel prediktor komponen parametrik dan satu variabel prediktor komponen nonparametrik dengan menggunakan satu titik knot dilakukan dengan menggunakan metode *Generalized Cross Validation* (GCV). Lokasi titik knot optimal pada model regresi semiparametrik dengan komponen paremetrik linier dan kuadratik ditunjukkan pada Tabel 4.2. Hasil nilai GCV minimum telah diurutkan dari nilai terkecil hingga terbesar berdasarkan model regresi semiparametrik birespon *spline truncated* komponen parametrik linier dan kuadratik dengan menggunakan *software* R.

Tabel 4.2 Nilai GCV yang Telah Diurutkan

Komponen Parametrik Linier		**Komponen Parametrik Kuadratik**	
Lokasi Knot	Nilai GCV	Lokasi Knot	GCV
5.641	1.605	2.961	1.493
5.258	1.611	5.641	1.495
6.024	1.614	3.727	1.496
1.812	1.622	5.258	1.497
6.407	1.622	6.024	1.501
4.875	1.623	4.110	1.502
4.492	1.631	1.812	1.502
4.110	1.638	4.875	1.503
2.195	1.640	4.492	1.503
3.727	1.642	3.344	1.503
2.961	1.645	2.578	1.504
3.344	1.645	2.195	1.504
2.578	1.649	6.407	1.506

Berdasarkan Tabel 4.2, maka untuk komponen parametrik linier diperoleh nilai GCV terkecil sebesar 1.605 dengan lokasi titik knot optimum pada respon satu (AHH) dan respon dua (Pengeluaran per Kapita) yaitu $K_{111} = 5.641$ dan $K_{211} = 5.641$. Untuk komponen parametrik kuadratik diperoleh nilai GCV terkecil sebesar 1.493 dengan lokasi titik knot optimum pada respon satu (AHH) dan respon dua (Pengeluaran per Kapita) yaitu K_{111}=2.961 dan K_{211}=2.961.

4.2.5 Pemodelan dengan Titik Knot Optimal

Pemodelan data AHH dan pengeluaran per kapita menggunakan regresi semiparametrik birespon *spline truncated* dengan komponen parametrik linier dilakukan berdasarkan persamaan (4.11) dan persamaan (4.12) sedangkan untuk model regresi semiparametrik birespon *spline truncated* dengan komponen kuadratik berdasarkan persamaan (4.13) dan (4.14). Hasil nilai estimasi parameter model regresi semiparametrik birespon komponen parametrik linier dan kuadratik ditunjukkan pada masing-masing Tabel 4.3 dan Tabel 4.4.

Tabel 4.3. Nilai Estimasi Parameter Komponen Parametrik Linier

Variabel	Respon 1		Respon 2	
	Parameter	Estimasi	Parameter	Estimasi
x_{0i}	$\hat{\beta}_{01}$	64.381	$\hat{\beta}_{02}$	2.461
x_{1i}	$\hat{\beta}_{11}$	0.958	$\hat{\beta}_{12}$	1.022
t_{1i}	$\hat{\alpha}_{11}$	-0.107	$\hat{\alpha}_{21}$	0.210
	$\hat{\alpha}_{111}$	-0.283	$\hat{\alpha}_{211}$	1.708

A. Estimasi dan interpretasi model untuk variabel komponen parametrik linier

Berdasarkan Tabel 4.3 model regresi semiparametrik birespon *spline truncated* komponen parametrik linier dapat dituliskan sebagai berikut.

Respon 1 (Angka Harapan Hidup)

$$\hat{y}_{1i} = 64.381 + 0.958x_{1i} - 0.107t_{1i} - 0.283(t_{1i} - K_{111})_+$$

Respon 2 (Pengeluaran per kapita)

$$\hat{y}_{2i} = 2.461 + 1.022x_{1i} + 0.201t_{1i} + 1.708(t_{1i} - K_{211})_+$$

Estimasi model untuk variabel prediktor nonparametrik TPT (t_{1i}) dengan mengasumsikan bahwa variabel prediktor rata-rata lama sekolah konstan, maka pengaruh TPT terhadap AHH ditunjukkan oleh fungsi *spline truncated* berikut.

$$\begin{aligned}
\hat{y}_{1i} &= -0.107t_{1i} - 0.283(t_{1i} - 5.641)_+ + c \\
&= \begin{cases} -0.107t_{1i} + c_1 & ; t_{1i} < 5.641 \\ -0.390t_{1i} + c_2 & ; t_{1i} \geq 5.641 \end{cases} \\
c_1 &= 64.381 + 0.958x_{1i} \\
c_2 &= 1.59 + c_1
\end{aligned} \tag{4.15}$$

Berdasarkan persamaan (4.15) diperoleh interpretasi estimasi model yaitu apabila TPT naik sebesar satu persen dan bernilai kurang dari 5.641 persen, maka AHH akan turun sebesar 0.107 persen. Apabila TPT naik sebesar satu persen dan bernilai lebih dari 5.641 persen maka AHH akan turun sebesar 0.390 persen.

Interpretasi model pengaruh variabel TPT terhadap pengeluaran perkapita dapat ditunjukkan pada persamaan (4.16) dengan mengasumsikan bahwa variabel prediktor rata-rata lama sekolah konstan.

$$\hat{y}_{2i} = 0.201 t_{1i} + 1.708 (t_{1i} - 5.641)_+ + c$$
$$= \begin{cases} 1.708 t_{1i} + c_1 & ; t_{1i} < 5.641 \\ 1.909 t_{1i} + c_2 & ; t_{1i} \geq 5.641 \end{cases}$$
$$c_1 = 2.641 + 1.022 x_{1i}$$
$$c_2 = 9.63 + c_1 \quad (4.16)$$

Berdasarkan persamaan (4.16) apabila TPT naik sebesar satu persen dan bernilai kurang dari 5.641 persen, maka pengeluaran per kapita akan naik sebesar 1.708 juta. Apabila TPT naik sebesar satu persen dan bernilai lebih dari 5.641 persen maka pengeluaran per kapita akan naik sebesar 1.909 juta.

Tabel 4.4. Nilai Estimasi Parameter Komponen Parametrik Kuadratik

Variabel	**Respon 1**		**Respon 2**	
	Parameter	Estimasi	Parameter	Estimasi
x_{0i}	$\hat{\theta}_{01}$	56.438	$\hat{\theta}_{02}$	9.730
x_{1i}	$\hat{\theta}_{11}$	3.519	$\hat{\theta}_{12}$	-1.143
x_{1i}^2	$\hat{\theta}_{21}$	-0.162	$\hat{\theta}_{22}$	0.138
t_{1i}	$\hat{\varphi}_{11}$	-0.779	$\hat{\varphi}_{11}$	0.487
	$\hat{\varphi}_{111}$	0.821	$\hat{\varphi}_{111}$	-0.137

B. Estimasi dan interpretasi model untuk variabel komponen parametrik kuadratik

Berdasarkan Tabel 4.4 model regresi semiparametrik birespon *spline truncated* komponen parametrik kuadratik dapat dituliskan sebagai berikut.

Respon 1 (Angka Harapan Hidup)

$$\hat{y}_{1i} = 56.438 + 3.519 x_{1i} - 0.162 x_{1i}^2 - 0.779 t_{1i} + 0.821 (t_{1i} - K_{111})_+$$

Respon 2 (Pengeluaran per kapita)

$$\hat{y}_{2i} = 9.730 - 1.143 x_{1i} + 0.138 x_{1i}^2 + 0.487 t_{1i} - 0.137 (t_{1i} - K_{211})_+$$

Estimasi model untuk variabel prediktor nonparametrik TPT (t_{1i}) dengan komponen paremetrik kuadratik, dan mengasumsikan bahwa variabel prediktor rata-rata lama sekolah konstan, maka pengaruh TPT terhadap AHH ditunjukkan oleh fungsi *spline truncated* berikut:

$$
\begin{aligned}
\hat{y}_{1i} &= -0.779t_{1i} + 0.821\left(t_{1i} - 2.961\right)_{+} + c \\
&= \begin{cases} -0.779t_{1i} + c_1 & ;t_{1i} < 2.961 \\ 0.042t_{1i} + c_2 & ;t_{1i} \geq 2.961 \end{cases}
\end{aligned}
$$

$$
c_1 = 56.438 + 3.519x_{1i}
$$

$$
c_2 = 0.124 + c_1 \tag{4.17}
$$

Berdasarkan persamaan (4.17) diperoleh interpretasi estimasi model yaitu apabila TPT naik sebesar satu persen dan bernilai kurang dari 2.961 persen, maka AHH akan turun sebesar 0.779 persen. Apabila TPT naik sebesar satu persen dan bernilai lebih dari 2.961 persen maka AHH akan naik sebesar 0.042 persen.

Interpretasi model pengaruh variabel TPT terhadap pengeluaran perkapita dapat ditunjukkan pada persamaan (4.18) dengan mengasumsikan bahwa variabel prediktor rata-rata lama sekolah konstan.

$$
\begin{aligned}
\hat{y}_{2i} &= 0.487t_{1i} - 0.137\left(t_{1i} - 2.961\right)_{+} + c \\
&= \begin{cases} 0.487t_{1i} + c_1 & ;t_{1i} < 2.961 \\ 0.350t_{1i} + c_2 & ;t_{1i} \geq 2.961 \end{cases}
\end{aligned}
$$

$$
c_1 = 9.730 - 1.143x_{1i} + 0.138x_{1i}^{\ 2}
$$

$$
c_2 = 0.405 + c_1 \tag{4.18}
$$

Berdasarkan persamaan (4.18) apabila TPT naik sebesar satu persen dan bernilai kurang dari 2.961 persen, maka pengeluaran per kapita akan naik sebesar 0.487 juta. Apabila TPT naik sebesar satu persen dan bernilai lebih dari 5.641 persen maka pengeluaran per kapita akan naik sebesar 0.350 juta.

Nilai GCV untuk model regresi semiparametrik birespon *spline truncated* dengan komponen kuadratik memiliki nilai yang lebih kecil apabila dibandingkan dengan model regresi semiparametrik birespon *spline truncated* dengan komponen linier. Hal ini dapat disebabkan karena adanya perbedaan struktur dari matriks variabel prediktor pada komponen parametrik linier dan kuadratik pada model. Hasil koefisien determinasi untuk komponen parametrik kuadratik dan komponen parametrik linier masing-masing sebesar 99.85% dan 99.87%. Artinya model yang telah dibentuk dengan dua data variabel prediktor yaitu rata-rata lama sekolah dan tingkat pengganguran terbuka dapat menjelaskan variabel respon AHH dan pengeluaran perkapita sebesar 99.85% dan 99.87%.

(Halaman ini sengaja dikosongkan)

BAB 5
KESIMPULAN DAN SARAN

Pada bagian ini akan dibahas mengenai kesimpulan dan saran dari hasil mengkaji estimator parameter model regresi semiparametrik birespon linier dan kuadratik menggunakan *spline truncated* dan memodelkan data angka harapan hidup dan pengeluaran perkapita di Provinsi Jawa Timur.

5.1 Kesimpulan

Kesimpulan dari hasil penelitian yang telah dilakukan adalah sebagai berikut:

1. Estimasi parameter model regresi semiparametrik birespon kuadratik menggunakan *spline truncated.*

$$\begin{aligned}\hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\boldsymbol{\theta}} + \boldsymbol{T}\hat{\boldsymbol{\varphi}} \\ &= \boldsymbol{XC(K)y} + \boldsymbol{TD(K)y} \\ &= \left(\boldsymbol{XC(K)} + \boldsymbol{TD(K)}\right)\boldsymbol{y} \\ &= \boldsymbol{A(K)y}\end{aligned}$$

dengan

$$\begin{aligned}\boldsymbol{A(K)} = &\left(\mathbf{M}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\left\{\boldsymbol{X}^T - \boldsymbol{X}^T\boldsymbol{WT}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\boldsymbol{T}^T\right\}\boldsymbol{W}\right) + \\ &\left(\mathbf{N}\left(\boldsymbol{T}^T\boldsymbol{WT}\right)^{-1}\left\{\boldsymbol{T}^T - \boldsymbol{T}^T\boldsymbol{WX}\left(\boldsymbol{X}^T\boldsymbol{WX}\right)^{-1}\boldsymbol{X}^T\right\}\boldsymbol{W}\right)\end{aligned}$$

dimana $\boldsymbol{A(K)}$ merupakan matriks yang berisi variabel prediktor dan titik knot.

Perbedaan estimasi parameter model regresi semiparametrik birespon menggunakan spline *truncated* dengan komponen parametrik kuadratik dan parametrik linier yaitu terletak pada struktur matriks $\boldsymbol{X}$ atau penyusun matriks variabel prediktor komponen parametrik.

2. Aplikasi pada data angka harapan hidup dan pengeluaran perkapita di Jawa Timur mendapatkan hasil sebagai berikut.
   a. Nilai GCV untuk model regresi semiparametrik birespon dengan komponen parametrik kuadratik lebih kecil jika dibandingkan dengan nilai GCV untuk model regresi semiparametrik birespon dengan komponen parametrik linier, yaitu 1.493 dan 1.605.
   b. Koefisien determinasi untuk model dengan komponen parametrik linier dan komponen parametrik kuadratik masing-masing sebesar 99.85% dan 99.87%.

c. Estimasi model regresi semiparametrik birespon menggunakan *spline truncated* dengan komponen parametrik linier menghasilkan model yang lebih baik dibandingkan dengan komponen paremetrik kuadratik, hal ini berdasarkan nilai GCV dan nilai koefisien determinasi yang dihasilkan memiliki selisih yang kecil sehingga diperoleh parsimoni model.

5.2 Saran

Berdasarkan hasil dan pembahasan, saran yang dapat diberikan adalah:

Penelitian ini terbatas pada metode yang digunakan yaitu *spline truncated* linier dengan komponen parametrik kuadratik, sehingga pada penelitian selanjutnya dapat dikembangkan menggunakan model regresi semiparametrik birespon *spline truncated* kuadratik dengan komponen parametrik kuadratik.

DAFTAR PUSTAKA

Afifah, N., Budiantara, I.N., dan Latra, I.N. (2017). Estimator of Kernel and Fourier Series in Semiparametric Regression. *International Conference on Mathematics : Education, Theory and Application.*

Amelia, D. (2014). *Model Regresi Nonparametrik Multirespon Spline Truncated untuk Data Longitudinal.* Tesis. Surabaya, Jurusan Statistika, Fakultas Matematika dan Sains Data, Institut Teknologi Sepuluh Nopember.

Anggreni, N.P.R, Suciptawati, N.L.P, dan Srinadi, I.G.A.M. (2018). Model Regresi Nonparametrik Spline Truncated Pada Jumlah Kasus Tuberkulosis Di Provinsi Bali Tahun 2016. *E-Jurnal Matematika,* **Vol.7**, No.3, 211-218.

Ardianti, A.V., Wibisono, S., dan Jumiati, A. (2015). *Faktor-faktor yang Mempengaruhi Angka Harapan Hidup di Kabupaten Jember*. Artikel Ilmiah Mahasiswa 2015. Universitas Jember.

Azizah, Z. (2018). *Interval Konfidensi untuk Parameter Model Regresi Semiparametrik Birespon dengan Pendekatan Spline Truncated.* Tesis. Surabya, Jurusan Statistika, Fakultas Matematika dan Sains Data, Institut Teknologi Sepuluh Nopember.

Badan Pusat Statistik, (2019). *Indeks Pembangunan Manusia* 2018. Jakarta.

Badan Pusat Statistik, (2019). *Indeks Pembangunan Manusia Jawa Timur* 2018. Jawa Timur.

Budiantara, I.N. (2005). Regresi Spline Linear. Makalah Seminar Nasional Matematika. Semarang: Jurusan Matematika, FMIPA Universitas Diponegoro (UNDIP).

Budiantara, I.N. (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 Purnomo, J.D.T. (2010). Mathematics XV National Conf. Manado: Manado University.

Budiantara, I. N. (2014). Pemodelan Regresi Nonparametrik dan Semiparametrik Spline (Konsep, Metode, dan Aplikasinya). *Prosiding Seminar Nasional Matematika.* Universitas Udayana. Denpasar. 1-16.

Budiantara, I. N. (2019). Regresi Nonparametrik Spline Truncated, Surabaya: ITS Press.

Drapper, N.R., dan Smith, H. (1992), *Applied Regression Analysis 2nd Edition*, Marcel Dekker, New York.

Eubank, R.L. (1999), *Spline Smoothing and Nonparametric Regression 2^{nd} Edition*, Marcel Dekker, New York.

Fitriani, R.N. (2015). *Analisis Faktor-Faktor yang Mempengaruhi Daya Beli Masyarakat di Jawa Barat*. Tugas Akhir. Universitas Pasundan.

Fitriyani, N. (2014), *Metode Cross Validation dan Generalized Cross Validation dalam Regresi Nonparametrik Spline (Studi Kasus Data Fertilitas di Jawa Timur)*. Tesis. Surabaya, Jurusan Statistika, Fakultas Matematika dan Sains Data, Institut Teknologi Sepuluh Nopember.

Gujarati, D. N., dan Porter, D. C. (2015). *Dasar-Dasar Ekonometrika Volume 1, 2^{nd} Edition*. Jakarta: Penerbit Salemba Empat.

Hardle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

Hidayati, L., Chamidah, N., dan Budiantara, I. N. (2019). *Spline Truncated Estimator in Multiresponse Semiparametric Regression Model for Computer based National Exam in West Nusa Tenggara. IOP Conf. Series: Materials Science and Engineering.* **Vol.546**, No.05, 20-29.

Mardianto, M.F.F., (2015). *Model Regresi Semiparametrik Birespon Dengan Pendekatan Deret Fourier.* Tesis, Surabaya, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember.

Mauly, A.V. (2014). *Faktor-faktor yang Mempengaruhi Indeks Kesehatan Kabupaten dan Kota di Provinsi Jawa Timur*. Tugas Akhir. Institut Teknologi Sepuluh Nopember, Surabaya.

Nurdiani, N., Herrhyanto, N., dan Dasari, D. (2017). Regresi Nonparametrik Birespon Spline. *Eureka Matika*. **Vol.5,** No.1, 106-121.

Pertiwi, R. (2012). *Permodelan Pengeluaran Per kapita per Kabupaten/Kota di Kalimantan Barat Menggunakan Metode Hirarki Bayesian*. Tesis. Surabaya, Jurusan Statistika, Institut Teknologi Sepuluh Nopember.

Putra, I.M.B., Srinadi, I.G.A.M., dan Sumarjaya, I.W. (2015). Pemodelan Regresi Spline (Studi Kasus: Herpindo Jaya Cabang Ngaliyan). *E-Jurnal Matematika*. **Vol.4**, No.3, 110-114.

Pratiwi, D.A., Budiantara, I.N., dan Wibowo, W. (2017). Pendekatan Regresi Semiparametrik Spline untuk Memodelkan Rata-Rata Umur Kawin Pertama di Provinsi Jawa Timur. *Jurnal sains dan seni ITS*. **Vol.6**, No.1, 129-136.

Prawanti, D.D., Budiantara, I.N., dan Purnomo, J.D.T. (2019). *Parameter Interval Estimation of Semiparametric Spline Truncated Regression Model for Longitudinal Data. IOP Conf. Series: Materials Science and Engineering.* **Vol.546**, No.05, 20-53.

Rencher, A.C, dan Schaalje, G.B. (2008). *Linear Model in Statistics,* 2^{nd} edition, America.

Ruppert, D., Wand, M.P., dan Carrol, R.J. (2003). *Semiparametric Regression*. Cambridge series in statistical and probabilistics mathematics. Cambridge: Univ. Press.

Simanjuntak, P. (2000). Pengantar Ekonomi Sumber Daya Manusia. Edisi kedua. Jakarta: Fakultas Ekonomi Universitas Indonesia.

Sugiantari, A.P, dan Budiantara, I.N. 2013. Analisis Faktor-faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline. *Jurnal Sains Dan Seni Pomits.* **Vol.2**, No.1, 37-41.

Sukirno, S. (2000). Makroekonomi Modern Perkembangan Pemikiran Dari Klasik Hingga Keynesian Baru. Jakarta: PT Raja Grafindo Persada.

Todaro, M.P. and Smith, S.C. (2012). *Economic Development Eleventh Edition*. Pearson. United States of America.

Wahba, G. (1990). *Spline Model for Observational Data*. SIAM Pensylvania.

Wang Y., Guo W. dan Brown, M.B. (2000). Smoothing Spline For Bivariate Data With Application To Association Between Hormones. *Statistica Sinica*. **Vol.10**, 377-397.

Wulandari, I.D.A.M.I dan Budiantara, I.N. (2014). Analisis Faktor-Faktor yang Mempengaruhi Persentase Penduduk Miskin dan Pengeluaran Perkapita Makanan di Jawa Timur Menggunakan Regresi Nonparametrik Birespon Spline. *Jurnal sains dan seni ITS.* **Vol.3**, No.1, 30-35.

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1. Data AHH, Pengeluaran Perkapita, Rata-Rata Lama Sekolah, dan TPT

Y1	Y2	X1	X2
71.52	8.527	7.19	1.43
72.43	9.426	7.17	3.87
73.35	9.4	7.27	4.17
73.74	10.455	8.06	2.61
73.16	10.327	7.27	3.37
72.37	10.853	7.68	4.25
72.26	9.844	7.18	3.24
69.7	8.931	6.21	2.55
68.74	9.09	6.07	4.09
70.34	11.828	7.12	3.67
66.27	10.429	5.62	3.9
68.73	9.692	6.11	1.92
66.71	10.7	5.71	4.15
70.01	9.933	6.83	6.11
73.82	14.168	10.24	4.73
72.24	12.454	8.18	4.27
72.04	10.999	8.21	4.64
71.25	11.768	7.61	2.64
70.97	11.351	7.57	3.81
72.3	11.539	7.95	3.92
71.92	11.187	6.88	3.83
71.07	9.926	6.77	4.19
71.01	10.048	6.52	2.83
72.04	11.108	7.83	3.17
72.46	12.845	8.96	5.82
69.94	8.393	5.33	5.25
67.79	8.569	4.36	2.41
67.22	8.536	6.35	2.92
70.94	8.722	5.23	1.79
73.8	11.976	9.91	3.63
73.36	13.391	9.9	4.06
72.93	16.158	10.16	6.79
70	11.796	8.49	3.64
71.18	12.931	9.1	4.55
73.01	13.155	9.99	2.45
72.59	15.616	11.11	3.85
73.98	17.157	10.46	6.12
72.37	12.466	8.77	3.12

Lampiran 2. *Syntax* R Estimasi Model Regresi Semiparametrik Komponen Parametrik Linier menggunakan Metode *Spline Truncated*

```
library(pracma)
library(MASS)
data = read.csv('D:/Data Asti_1LP_bagi.csv',sep=';',header=TRUE)

#definisi variabel
data
y1<-as.matrix(data[,1])
y2<-as.matrix(data[,2])
y<-rbind(y1,y2)
x1<-as.matrix(data[,c(3:4)])
x2<-as.matrix(data[,c(3:4)])
x<-rbind(x1,x2)
nt<-nrow(x);nt
t<-2;t
n<-nt/t;n
p<-ncol(x);p                    #banyaknya variabel prediktor
P<-1                            #komponen Nonparametrik
Q<-1                            #komponen parametrik
r<-1      ;r                    #banyaknya titik knot
q<-1      ;q                    #derajat polinomial

#========matrix W=============
a=var(y1)
b=var(y2)
W11=diag(rep(a,n))
W12=matrix(0,n,n)
W22=diag(rep(b,n))
W=cbind(rbind(W11,W12),rbind(W12,W22))

#====matrix Y=======
Yk<-rbind(y1,y2)
Yk<-as.vector(Yk)

#====matrix X=====
Xt0<-matrix(1,nt,1)
Xt<-as.matrix(cbind(Xt0,x))

#=====knot======
nknot=15
nkomb=nknot
knot1=matrix(0,nkomb,1)
a=1
for (l in 1:t)
{
 xx=as.matrix(Xt[a:(l*n),3])
 kn=matrix(seq(min(xx[,1]),max(xx[,1]),length.out=nknot),nknot,1)
 knot=matrix(0,nkomb,1)
 v=1
 for (i in 1:nknot)
 {
  knot[v,]=cbind(kn[i,1])
  v=v+1
 }
```

```
knot1=cbind(knot1,knot)
  a=a+t
}
knot1=as.matrix(knot1[2:(nrow(knot1)-1),2:ncol(knot1)])
nknot=nrow(knot1)
knot1

#========GCV minimum========
MSE=matrix(0,nrow=nknot)
GCV=matrix(0,nrow=nknot)
Z=matrix(0,nt,(((r+q)*Q+P+1)*t))
for (i in 1:nknot)
{
  b=1
  a=1
  Xk=matrix(0,1,r*1)
  for (j in 1:t)
  {
    xa=matrix(0,n)
    aa=cbind(pmax(0,Xt[b:(j*n),3]-knot1[i,a]))
    xb=as.matrix(cbind(xa,aa))
    xc=as.matrix(xb[,2])
    Xk=rbind(Xk,xc)
    b=b+n
    a=a+1
  }
Xk=Xk[2:nrow(Xk),]
  Xk=cbind(Xt,Xk)
  a=1
  rq=(r+q)*Q+P+1
  b=1
  for (j in 1:t)
  {
    Z[a:(j*n),b:(j*rq)]=Xk[a:(j*n),]
    a=a+n
    b=b+rq
  }
  B=pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W%*%Yk
  Ak=Z%*%pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W
  ID=diag(1,nt,nt)
  AK=ID-Ak
  yhat=Ak%*%Yk
  error=Yk-yhat
  MSE[i]=((t(error)%*%error)/nt)
  db=(sum(diag(AK))/nt)
  GCV[i]=MSE[i]/((db)^2)
}
optimum=cbind(knot1,MSE,GCV)
GCVmin=optimum[order(optimum[,ncol(optimum)]),]
GCVmin[1,]
```

```
#=======validasi optimum======
#matriks z

Z=matrix(0,nt,(((r+q)*Q+P+1)*t))
Z
Xk=matrix(0,1,r*1)
b=1
a=1
for(j in 1:t)
{
  xa=matrix(0,n)
  aa=cbind(pmax(0,Xt[b:(j*n),3]-GCVmin[1,a]))
  xb=as.matrix(cbind(xa,aa))
  xc=as.matrix(xb[,2])
  Xk=rbind(Xk,xc)
  b=b+n
  a=a+1
}
Xk=Xk[2:nrow(Xk),]
Xk=cbind(Xt,Xk)
a=1
rq=(r+q)*Q+P+1
b=1
for(j in 1:t)
{
  Z[a:(j*n),b:(j*rq)]=Xk[a:(j*n),]
  a=a+n
  b=b+rq
}
B=pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W%*%Yk
Ak=Z%*%pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W
yhat=Ak%*%Yk
SSE=sum((Yk-yhat)^2)
SSR=sum((yhat-mean(Yk))^2)
SST=sum((Yk-mean(Yk))^2)
MSE=(SSE/(76-8))
MSR=SSR/(76-7)
Rsq=(SSR/(SSR+SSE))*100
n1<-nrow(B)
error<-yhat-data[,1]
```

Lampiran 3. *Syntax* R Estimasi Model Regresi Semiparametrik Komponen Parametrik Kuadratik menggunakan Metode *Spline Truncated*

```
library(pracma)
library(MASS)
data = read.csv('D:/Data Asti_1KP_bagi.csv',sep=';',header=TRUE)

#definisi variabel
data
y1<-as.matrix(data[,1])
y2<-as.matrix(data[,2])
y<-rbind(y1,y2)
x1<-as.matrix(data[,c(3:4)])
x2<-as.matrix(data[,c(3:4)])
x<-rbind(x1,x2)
nt<-nrow(x);nt
t<-2;t
n<-nt/t;n
p<-ncol(x);p                    #banyaknya variabel prediktor
P<-1                            #komponen Nonparametrik
Q<-2                            #komponen parametrik
r<-1      ;r                    #banyaknya titik knot
q<-1      ;q                    #derajat polinomial

#========matrix W==============
a=var(y1)
b=var(y2)
W11=diag(rep(a,n))
W12=matrix(0,n,n)
W22=diag(rep(b,n))
W=cbind(rbind(W11,W12),rbind(W12,W22))

#====matrix Y=======
Yk<-rbind(y1,y2)
Yk<-as.vector(Yk)

#====matrix X=====
Xt0<-matrix(1,nt,1)
Xt<-as.matrix(cbind(Xt0,x))

#=====knot=======
nknot=15
nkomb=nknot
knot1=matrix(0,nkomb,1)
a=1
for (l in 1:t)
{
  xx=as.matrix(Xt[a:(l*n),3])
  kn=matrix(seq(min(xx[,1]),max(xx[,1]),length.out=nknot),nknot,1)
  knot=matrix(0,nkomb,1)
  v=1
  for (i in 1:nknot)
  {
    knot[v,]=cbind(kn[i,1])
    v=v+1
  }
```

```
knot1=cbind(knot1,knot)
  a=a+t
}
knot1=as.matrix(knot1[2:(nrow(knot1)-1),2:ncol(knot1)])
nknot=nrow(knot1)
knot1

#========GCV minimum========
MSE=matrix(0,nrow=nknot)
GCV=matrix(0,nrow=nknot)
Z=matrix(0,nt,(((r+q)*Q+P+1)*t))
for (i in 1:nknot)
{
  b=1
  a=1
  Xk=matrix(0,1,r*1)
  for (j in 1:t)
  {
    xa=matrix(0,n)
    aa=cbind(pmax(0,Xt[b:(j*n),3]-knot1[i,a]))
    xb=as.matrix(cbind(xa,aa))
    xc=as.matrix(xb[,2])
    Xk=rbind(Xk,xc)
    b=b+n
    a=a+1
  }
Xk=Xk[2:nrow(Xk),]
  Xk=cbind(Xt,Xk)
  a=1
  rq=(r+q)*Q+P+1
  b=1
  for (j in 1:t)
  {
    Z[a:(j*n),b:(j*rq)]=Xk[a:(j*n),]
    a=a+n
    b=b+rq
  }
  B=pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W%*%Yk
  Ak=Z%*%pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W
  ID=diag(1,nt,nt)
  AK=ID-Ak
  yhat=Ak%*%Yk
  error=Yk-yhat
  MSE[i]=((t(error)%*%error)/nt)
  db=(sum(diag(AK))/nt)
  GCV[i]=MSE[i]/((db)^2)
}
optimum=cbind(knot1,MSE,GCV)
GCVmin=optimum[order(optimum[,ncol(optimum)]),]
GCVmin[1,]
```

```
#=======validasi optimum========
#matriks z
Z=matrix(0,nt,(((r+q)*P+Q+1)*t))
Xk=matrix(0,1,r*1)
b=1
a=1
for(j in 1:t)
{
  xa=matrix(0,n)
  aa=cbind(pmax(0,Xt[b:(j*n),4]-GCVmin[1,a]))
  xb=as.matrix(cbind(xa,aa))
  xc=as.matrix(xb[,2])
  Xk=rbind(Xk,xc)
  b=b+n
  a=a+1
}
Xk=Xk[2:nrow(Xk),]
Xk=cbind(Xt,Xk)
a=1
rq=(r+q)*P+Q+1
b=1
for(j in 1:t)
{
  Z[a:(j*n),b:(j*rq)]=Xk[a:(j*n),]
  a=a+n
  b=b+rq
}
B=pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W%*%Yk
Ak=Z%*%pinv(t(Z)%*%W%*%Z)%*%t(Z)%*%W
yhat=Ak%*%Yk
SSE=sum((Yk-yhat)^2)
SSR=sum((yhat-mean(Yk))^2)
SST=sum((Yk-mean(Yk))^2)
MSE=(SSE/(76-10))
MSR=SSR/(76-9)
Rsq=(SSR/(SSR+SSE))*100
n1<-nrow(B)
error<-yhat-data[,1]
```

Lampiran 4. Uji Korelasi Pearson

Correlations

		Y1_AHH	Y2_PKPT
Y1_AHH	Pearson Correlation	1	.527**
	Sig. (2-tailed)		.001
	N	38	38
Y2_PKPT	Pearson Correlation	.527**	1
	Sig. (2-tailed)	.001	
	N	38	38

**. Correlation is significant at the 0.01 level (2-tailed).

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, Mahasiswa Departemen Statistika Fakultas Sains dan Analitika Data (FSAD) ITS:

Nama : Adawiyah Asti Khalil

NRP : 06211850010002

Program Studi : Magister Statistika

menyatakan bahwa data yang digunakan dalam Tesis / ~~Disertasi~~ ini merupakan data sekunder yang diambil dari ~~penelitian~~ / ~~buku~~ / ~~Tesis~~ / ~~Disertasi~~ / publikasi lainnya yaitu:

Sumber : Website Badan Pusat Statistik Provinsi Jawa Timur, https://jatim.bps.go.id/

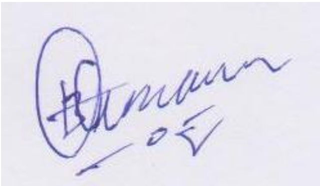
Keterangan : Data tahun 2018

Surat Pernyataan ini dibuat dengan sebenarnya, apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan berlaku.

Surabaya, Desember 2019

Mengetahui / Menyetujui :

Pembimbing Tesis



Prof. Dr. Drs. I Nyoman Budiantara, M.Si
NIP.19650603 198903 1 003

Mahasiswa

Adawiyah Asti Khalil

*Coret yang tidak perlu

(Halaman ini sengaja dikosongkan)

BIOGRAFI PENULIS

Adawiyah Asti Khalil lahir di Surabaya pada tanggal 24 Februari 1995. Penulis merupakan anak pertama dari pasangan Bapak Aba'ad dan Ibu Suhati. Penulis memulai jenjang pendidikan formal di Sekolah Dasar Negeri 3 Ungga, Kabupaten Lombok Tengah pada tahun 2001-2007. Kemudian penulis melanjutkan pendidikan Sekolah Menengah Pertama Negeri 1 Praya Barat Daya tahun 2007-2010. Penulis melanjutkan pendidikan ke jenjang berikutnya yaitu Sekolah Menengah Atas Negeri 1 Praya, Kabupaten Lombok Tengah pada tahun 2010-2013. Selanjutnya, penulis menempuh pendidikan Sarjana (S1) program studi Matematika, FMIPA, Universitas Mataram, NTB dan berhasil menyelesaikan pendidikan S1 pada tahun 2017. Tahun 2018 penulis melanjutkan studi Pascasarjana (S2) di Departemen Statistika, Fakultas Sains dan Analitika Data (FSAD), Institut Teknologi Sepuluh Nopember, Surabaya. Segala saran dan kritik yang membangun, sangat penulis harapkan untuk evaluasi dan kebaikan ke depannya. Penulis dapat dihubungi melalui email adawiyah.asti24@gmail.com.

(Halaman ini sengaja dikosongkan)