ESTIMASI INTERVAL UNTUK KURVA REGRESI CAMPURAN SPLINE TRUNCATED DAN KERNEL DALAM REGRESI NONPARAMETRIK

SYISLIAWATI
NRP. 1314201214

DOSEN PEMBIMBING :
Dr. Wahyu Wibowo, S.Si, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

STATISTICAL INFERENCE FOR CURVE OF NONPARAMETRIK REGRESSION WITH COMBINATION OF SPLINE AND KERNEL

SYISLIAWATI
NRP. 1314201214

SUPERVISOR :
Dr. Wahyu Wibowo, S.Si, M.Si
Prof. Dr. Drs. I Nyoman Budiantara, M.Si

MAGISTER PROGRAM
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

ESTIMASI INTERVAL UNTUK PARAMETER REGRESI CAMPURAN *SPLINE TRUNCATED* DAN KERNEL DALAM REGRESI NONPARAMETRIK

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

SYISLIAWATI
NRP. 1314 201 214

Tanggal Ujian : 20 Juli 2016
Periode Wisuda : September 2016

Disetujui oleh :

1. **Dr. Wahyu Wibowo, S.Si., M.Si.** **(Pembimbing I)**
   NIP. 19740328 199802 1 001

2. **Prof. Dr. Drs. I Nyoman Budiantara, M.Si.** **(Pembimbing II)**
   NIP. 19650603 198903 1 003

3. **Dr. Vita Ratnasari, S.Si., M.Si.** **(Penguji)**
   NIP. 19700910 199702 2 001

4. **Dr. Kartika Fitriasari, M.Si.** **(Penguji)**
   NIP. 19691212 199303 2 002

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
PROGRAM PASCASARJANA

Direktur Program Pascasarjana,

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.
NIP. 196012021987011001

ESTIMASI INTERVAL UNTUK KURVA REGRESI CAMPURAN SPLINE TRUNCATED DAN KERNEL DALAM REGRESI NONPARAMETRIK

Nama Mahasiswa	: Syisliawati
NRP	: 1314201214
Pembimbing 1	:Dr. Wahyu Wibowo, M.Si.
Pembimbing 2	: Prof. Dr. Drs. I Nyoman Budiantara, M.Si

ABSTRAK

Salah satu bagian penting dalam ilmu statistika adalah persoalan inferensi yaitu penarikan kesimpulan secara statistik. Distribusi populasi atau distribusi variabel acak yang digunakan pada teknik inferensi ini mempunyai bentuk matematik yang diketahui, akan tetapi memuat beberapa parameter yang tidak diketahui. Pada umumnya analisis regresi digunakan untuk menyelidiki adanya hubungan antara variabel respon dan variabel prediktor dengan mengamati pola kecenderungan pola hubungannya. Apabila kurva regresi tidak diketahui maka dipergunakan metode regresi nonparametrik. Penelitian ini bertujuan menurunkan bentuk estimator dan interval konfidensi terpendek untuk angka kematian bayi Propinsi Jawa Timur menggunakan estimator campuran *kernel* dan *spline* pada regresi nonparametrik. Angka Kematian Bayi (AKB) di Indonesia pada tahun 2011 adalah 34 per 1000 kelahiran. Angka ini masih tergolong tinggi karena maksimal AKB menurut *Millennium Development Goals* (*MDGs*) adalah 23 per 1000 kelahiran. Beberapa provinsi di Indonesia AKB juga masih tergolong tinggi, seperti Propinsi Jawa Timur yang belum mencapai target dari *MDGs*. Hal ini mengisyaratkan pemerintah propinsi untuk mengetahui faktor faktor apa saja yang mempengaruhi dari tingginya AKB tersebut dan segera membuat program untuk mengurangi terjadinya AKB. Adapun variabel-variabel yang diduga berpengaruh yaitu persentase persalinan yang dilakukan dengan bantuan non medis, persentase wanita perkawinan pertama kurang dari 17 tahun, rata-rata jumlah pengeluaran rumah tangga perkapita sebulan, persentase bayi usia 0-11 bulan yang diberi ASI dan jumlah sarana kesehatan. Berdasarkan hasil penelitian di peroleh model terbaik yaitu pada tiga titik knot dimana nilai MSE dan R^2 yang terbaik berturut-turut adalah 4,11 dan 88,97% dan semua parameternya signifikan terhadap angka kematian bayi.

Kata kunci : *Regresi Nonparametrik Kernel dan Spline, GCV dan angka Kematian Bayi (AKB)*

Abstract

One of the important parts of statistical study is statistical inference. Statistical inference is the process of drawing conclusions. Population distribution or random variable distribution used in this inference techniques has the mathematical shapes known, but contain of some parameters unknown. Generally regression analysis used to investigate the connection between response variables and predictor variables are by observing the tendency of their relationship. When the regression curve unknown, used the nonparametric regression method. Nonparametric regression for examples already known include spline function and kernel. This research aims to reduce the shape of estimator and confidence intervals for the shortest infant mortality rate in East Java uses the combination of estimator between kernel and spline. The infant mortality rate in Indonesia in 2011 is 34 per 1000 live births. This figure still relatively high because the maximum infant mortality based on Millennium Development Goals (*MDGs*) is 23 per 1000 live births. The infant mortality rate in several provinces remain high, such as province of East Java not yet reached the target based on *MDGs*. Thus suggest the provincial government to determine what factors affecting the high of the infant mortality rate. The provincial government should make program to reduce the infant mortality rate. Variables allegedly influential such as the percentage of childbirth with the assistance of non medical, the percentage of the first marriage of woman less than 17 years old, the average expenditures of household per capita a month, the percentage of infants aged 0-11 months who were breastfed and numbers of health facilities.

Keywords : *kernel nonparametric regression* and *spline*, *GCV* and *infant mortality rate*.

DAFTAR ISI

	Hal
ABSTRAK……………………………………………………………....	i
DAFTAR ISI……………………………………………...................	iii
DAFTAR TABEL……………………………………………………......	v
DAFTAR GAMBAR………………………………………………….....	vi
KATA PENAGANTAR…………………………………………….….....	vii
BAB 1. PENDAHULUAN………………………………………….....	
1.1 Latar Belakang……………………………………………	1
1.2 Perumusan Masalah………………………………………	4
1.3 Tujuan Penelitian…………………………………………	4
1.4 Manfaat Penelitian……………………………………….	4
1.5 Batasan Permasalahan……………………………………	5
BAB 2. TINJAUAN PUSTAKA…………………………………...	7
2.1 Analisis Regresi …………………………………………	7
2.2 Pemodelan Regresi Parametrik dan Nonparametrik ……...	7
2.3 Spline dalam Regresi Nonparametrik ………………………	8
2.4. Regresi Nonparametrik Kernel……………………………….	9
2.5 Estimator Campuran Spline dan Kernel………………………	11
2.6 Koefisien Determinasi (R^2)………………………………....	13
2.7 Interval Konfidensi……………………………………………	13
BAB 3. METODE PENELITIAN……………………………………….	15
3.1 Sumber Data………………………………………………..	15
3.2 Variabel Penelitian……………………………………………	15
3.3 Metode Penelitian…………………………………………….	17
BAB 4. HASIL DAN PEMBAHASAN…………………………………..	19
4.1 Interval Konfidensi Terpendek untuk Parameter Model Komponen Spline Truncated Variansi diketahui……….......	19
4.2 Interval Konfidensi Terpendek untuk Parameter Model Komponen Spline Truncated Variansi tidak diketahui……..	23

Hal
4.3 Aplikasi pada Model Angka Kematian Bayi di Provinsi Jawa Timur………………………………………………………… 28
4.3.1 Analisis Dekriptif AKB di Provinsi Jawa Timur…. 28
4.3.2 Model AKB dengan Provinsi Jawa Timur……….. 30
4.3.3 Pemilihan Titik Knot Optimal pada Spline Linear.. 34
4.3.3.1 Pemilihan Titik Knot optimal Spline Truncated dan *kernel* Satu Titik Knot…………………… 34
4.3.3.2 Pemilihan Titik Knot Optimal Spline Truncated dan Kernel dengan Dua Titik Knot……………. 35
\ 4.3.3.3 Pemilihan Titik Knot Optimal Spline Truncated dan Kerne dengan Tiga Titik Knot…………….. 36
4.3.4 Interval Konfidensi Provinsi Jawa Timur……… 38
BAB 5………………………………………………………… 41
5.1 Kesimpulan……………………………………………………. 41
5.2 Saran………………………………………………………… 43
DAFTAR PUSTAKA 45
LAMPIRAN 47
BIOGRAFI PENULIS

DAFTAR TABEL

Halaman

Tabel 3.1	Variable Penelitian	15
Tabel 3.2	Struktur Data	16
Tabel 4.1	Statistika Deskriptif AKB dengan 5 Variabel Prediktor	30
Tabel 4.2	Skor GCV untuk 1 Titik Knot	35
Tabel 4.3	Skor GCV untuk 2 Titik Knot	36
Tabel 4.4	Skor GCV untuk 3 Titik Knot	37
Tabel 4.5	Perbandingan Nilai GCV Minimum dari Titik Knot	39
Tabel 4.6	Batas Atas dan Batas Bawah Interval Konfidensi	40

DAFTAR GAMBAR

Judul Gambar Halaman

Gambar 4.1 *Plot* antara AKB dengan persentase persalinan dengan bantuan nonmedis.. 30

Gambar 4.2 *Plot* Antara AKB dengan Persentase wanita perkawinan pertama < 17 Tahun.. 30

Gambar 4.3 *Plot* AKB dengan rata-rata jumlah pengeluaran rumah tangga perkapita sebulan 31

Gambar 4.4 *Plot* AKB dengan persentase bayi usia 0-11 bulan di beri ASI 32

Gambar 4.5 *Plot* AKB dengan Jumlah Sarana Kesehatan 33

DAFTAR LAMPIRAN

Halaman

Lampiran 1	Data yang digunakan..	49
Lampiran 2	Program Interval Konfidensi *Spline* 1 Titik Knot.......	51
Lampiran 3	Program Interval Konfidensi *Spline* 2 Titik Knot........	58
Lampiran 4	Program Interval Konfidensi *Spline* 3 Titik Knot.........	65

Halaman ini sengaja di kosongkan

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Kematian bayi adalah kematian yang terjadi pada bayi pada selang waktu antara saat setelah bayi lahir sampai dengan bayi belum berumur tepat satu tahun. Besaran yang menyatakan kemungkinan terjadinya bayi mati setelah lahir sampai mencapai usia satu tahun dalam perseribu kelahiran hidup dinamakan Angka Kematian Bayi (AKB). Berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI) 2012, AKB di Indonesia mencapai angka 32/1000 kelahiran hidup bayi. Hal ini dinyatakan masih cukup tinggi jika dibandingkan dengan standar dari *Millennium Development Goals* (MDGs) yaitu 23/1000 untuk AKB. Oleh karena itu, pemerintah diharapkan mampu menekan nilai AKB melalui program-program yang dicanangkan atau dengan mengetahui dan mengatasi faktor-faktor yang berpengaruh secara signifikan terhadap tingginya nilai AKB.

Penekanan nilai AKB hingga mencapai target dari MDGs atau bahkan lebih rendah dapat menggunakan konsep mengatasi faktor-faktor penyebab tingginya nilai AKB. Beberapa penelitian yang fokus pada hal ini antara lain Pramasita (2005), Jayanti (2007) dan Puspitasari (2012). Pramasita (2005) menyatakan bahwa proses persalinan dengan menggunakan tenaga non-medis dan banyaknya trauma lahir sangat berpengaruh terhadap nilai AKB. Lebih lanjut, Jayanti (2007) pada penelitiannya menyimpulkan bahwa bayi yang tidak diberi air susu ibu (ASI), wanita yang tidak pernah sekolah atau tidak tamat SD/MI, persalinan yang menggunakan tenaga non-medis, wanita yang berkeluarga di bawah umur 17 tahun dan penduduk golongan sosial ekonomi menengah ke bawah sangat berpengaruh pada tingginya nilai AKB. Berdasar pada penelitian yang dilakukan Puspitasari (2012), dengan menggunakan data SUSENAS 2012 oleh BPS dan data dari Dinas Kesehatan Propinsi Jawa Timur tahun 2011, menyatakan bahwa jumlah sarana medis, persentase persalinan dengan menggunakan tenaga non-medis, persentase ibu yang tidak melakukan kunjungan kehamilan dan persentase bayi yang tidak diberi ASI berpengaruh signifikan pada nilai AKB.

Data nilai AKB pada penelitian-penelitian tersebut, berdasarkan diagram pencarnya dengan masing-masing variabel prediktornya yang telah disebutkan mempunyai kecenderungan sebagian datanya memiliki pola yang beruah-ubah pada sub-sub interval tertentu dan sebagian tidak mengitu pola tertentu (linier atau polinomial berderajat tertentu) dan sehingga pendekatan model regresi yang sesuai adalah model regresi nonparametrik campuran spline dan kernel.

Dalam penelitian ini akan dijelaskan variabel- variabel yang berpengaruh terhadap angka kematian bayi (AKB). Untuk mengetahui variabel-variabel yang berpengaruh terhadap (AKB) ini dapat dilakukan pengujian hipotesis dan interval konvidensi dalam estimator campuran spline dan kernel. Tetapi pengujian hipotesis dalam estimator campuran spline dan kernel sejauh ini menurut sepengetahuan penulis belum ada yang di publikasikan atau belum ditemukan di dalam berbagai pustaka. Oleh karena itu dalam penelitian ini digunakan interval konfidensi untuk mengetahui variabel-variabel yang mempengaruhi angka kematian bayi (AKB).

Estimasi interval adalah estimasi pengembangan estimasi titik. Nilai taksiran parameter tidak terfokus pada satu titik tetapi di dasarkan pada range tertentu, sehingga estimasinya memiliki nilai tertinggi (max) dan nilai terendah (min). interval ini lebih dikenal dengan interval konfidensi. Nilai yang muncul adalah nilai yang didasarkan probabilitas tertentu, dalam praktek biasanya yang dipilih 90%, 95% atau 99%.

Sehingga dalam penelitian ini akan dilakukan estimasi parameter komponen spline regresi campuran *spline* dan *kernel* dalam regresi nonparametrik pada Angka Kematian Bayi (AKB) di provinsi jawa timur.

1.2 Rumusan Masalah

Berdasarkan uraian diatas maka, perumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana mendapatkan bentuk interval konfidensi untuk parameter-parameter pada model spline dalam estimator campuran spline dan kernel.
2. Bagaimana mengaplikasikan model estimator campuran spline dan kernel pada data Angka Kematian Bayi (AKB) di Provinsi Jawa Timur.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan diatas, maka tujuan dari penelitian ini adalah:

1. Mendapatkan bentuk interval konfidensi untuk parameter-parameter pada model spline dalam estimator campuran spline dan kernel.
2. Mengaplikasikan model estimator campuran spline dan kernel pada data Angka Kematian Bayi (AKB) di Provinsi Jawa Timur.

1.4 Manfaat Penelitian

1. Mengembangkan wawasan keilmuan dan pengetahuan tentang regresi nonparametrik *spline truncated*.
2. Mengembangkan wawasan keilmuan dan pengetahuan tentang regresi nonparametrik kernel.
3. Mengembangkan wawasan keilmuan dan pengetahuan tentang Estimator Campuran Kernel dan Spline dalam regresi nonparametrik.
4. Hasil penelitian diharapkan menjadi bahan masukan atau acuan dalam penelitian selanjutnya yang akan dilakukan.

1.5 Batasan Penelitian

Batasan penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data tahun 2011.
2. Model spline yang digunakan adalah spline truncated.
3. Estimator kernel yang digunakan adalah estimator kernel Naedaraya dan Watson.
4. Pemilihan titik knot dan bandwith optimal menggunakan metode GCV.
5. Dalam aplikasi digunakan spline truncated linear dengan satu, dua dan tiga knot.
6. Estimasi interval yang dilakukan dalam penelitian ini hanya untuk komponen spline dalam regresi nonparametrik campuran.

Halaman ini sengaja di kosongkan

BAB 2
TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi merupakan suatu analisis statistika yang digunakan untuk memodelkan pola hubungan antara variabel prediktor dan variabel respon. Terdapat dua pendekatan dalam model regresi, yaitu regresi parametrik dan regresi nonparametrik. Pendekatan regresi parametrik digunakan jika bentuk kurva regresi diketahui. Jika pola hubungan data membentuk pola linear maka digunakan pendekatan regresi parametrik linear. Jika pola hubungan data membentuk pola kuadrat maka digunakan pendekatan regresi kuadratik, dan lain-lain (Eubank,1988). Bentuk pola hubungan dapat diidentifikasi berdasarkan pada informasi masa lalu atau *scatter plot* data (Hardle, 1990).

2.2 Pemodelan Regresi Parametrik dan Nonparametrik

Permodelan regresi parametrik digunakan apabila kurva regresi data membentuk pola tertentu, seperti linier, kuadratik ataupun kubik. Model regresi parametrik dalam penggunaanya memerlukan informasi dari masa lalu atau sumber informasi lain yang tersedia yang dapat menggambarkan secara detail tentang data tersebut. Model regresi parametrik juga mempunyai asumsi yaitu bentuk kurva regresi harus diketahui (Eubank, 1999). Metode untuk mengestimasi parameter adalah metode *Least Square* dan *Maximum Likelihood* (Wahba, 1990). Pada umumnya, variabel respon y dapat dihubungkan dengan k variabel-variabel prediktor. Model tersebut dalam bentuk (Montgomery & Hines, 1972)

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i, i = 1, 2, ..., n \tag{2.1}$$

dimana y_i merupakan variabel respon, $x_1, x_2, ..., x_k$ sebagai variabel prediktor, ε_i merupakan error random independen berdistribusi normal dengan mean nol dan varians σ^2 parameter $\beta_0, \beta_1, \beta_2, ..., \beta_k$ tidak diketahui. Estimasi regresi parametrik sangat efisien jika model yang diasumsikan ini benar, tetapi jika tidak benar akan menyebabkan interpretasi data yang menyesatkan (Simonoff, 1996).

Secara umum bentuk regresi parametrik multivariabel dengan k variabel prediktor pada persamaan (2.1) dapat ditulis dalam bentuk matriks sebagai :

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, dimana (2.2)

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Pada umumnya, $\mathbf{y}$ adalah sebuah vektor (n x 1) dari observasi-observasi, $\mathbf{X}$ adalah sebuah matriks (n x (k+1)) dari variabel-variabel bebas, $\boldsymbol{\beta}$ adalah sebuah vektor ((k+1)x1) dari koefisien-koefisien regresi dan $\boldsymbol{\varepsilon}$ adalah sebuah vektor (nx1) dari error random. Parameter $\boldsymbol{\beta}$ diestimasi dengan metode kuadrat terkecil yang meminimumkan $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ dimana :

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Dengan menurunkan parsial $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ terhadap $\boldsymbol{\beta}$ dan menyamakan dengan nol, maka diperoleh estimator :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \qquad (2.3)$$

2.3 Spline dalam Regresi Nonparametrik

Spline dalam regresi nonparametrik mempunyai sifat fleksibilitas yang tinggi dan mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan (Eubank, 1988; Budiantara, 2006). Kemampuan mengestimasi perilaku data ini ditunjukkan oleh fungsi *truncated* (potongan-potongan) yang melekat pada estimator dan potongan-potongan tersebut yang disebut titik knot. Titik knot merupakan titik perpaduan bersama yang menunjukkan perubahan pola perilaku fungsi pada selang yang berbeda. Spline merupakan salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data. Fungsi spline berderajat

p adalah sebarang fungsi yang secara umum dapat disajikan dalam bentuk sebagai berikut :

$$f(x_i) = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{j=1}^{k} \beta_{j+p} (x_i - k_j)_+^p \qquad (2.5)$$

dengan β_j adalah konstanta riil, dan

$$(x_i - k_j)_+^p = \begin{matrix} (x_i - k_j)^p & ; x \geq k_j \\ 0 & ; x < k_j \end{matrix} \qquad (2.6)$$

Jika p = 1, 2, dan 3 diperoleh berturut-turut spline linear, spline kuadratik dan spline kubik serta $k_j$adalah titik knot.

Apabila diasumsikan *error* ε_i berdistribusi normal independen dengan rata-rata nol dan variansi σ^2, maka y_i pada model regresi juga berdistribusi normal dengan rata-rata $f(x_i)$ dan variansi σ^2. Estimasi untuk parameter β dengan menggunakan metode *least square*, yaitu dengan meminimumkan jumlah kuadrat *error* adalah sebagai berikut,

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_j - f(x_i)\right)^2$$

dengan penyajian matriks, diperoleh :

$$\sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta)$$

Estimator β dapat di peroleh dari meminimumkan

$$Q(\beta) = (y - X\beta)^T (y - X\beta)$$

2.4 Regresi Nonparametrik Kernel

Estimator kernel mempunyai kelebihan yaitu fleksibel, bentuk matematisnya mudah dan dapat mencapai tingkat kekonvergenan yang relative cepat. Kurva regresi $g(t_i)$ yang dihampiri fungsi kernel, estimasi kurva regresi dapat disajikan dalam bentuk :

$$\hat{g}_\alpha(t) = n^{-1} \sum_{i=1}^{n} \left[\frac{K_\alpha(t - t_i)}{n^{-1} \sum_{j=1}^{n} K_\alpha(t - t_j)} \right] y_i = n^{-1} \sum_{i=1}^{n} W_{\alpha i}(t) y_i \qquad (2.8)$$

dimana :

$$W_{\alpha i}\left(t\right)=\frac{K_{\alpha}\left(t-t_{i}\right)}{n^{-1}\sum_{j=1}^{n}K_{\alpha}\left(t-t_{j}\right)};\quad K_{\alpha}\left(t-t_{i}\right)=\frac{1}{\alpha}K\left(\frac{t-t_{i}}{\alpha}\right)$$

dengan K merupakan fungsi kernel. Menurut Hardle (1990), fungsi kernel K dapat berupa :

- Kernel Gaussian : $K(z)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^{2}}{2})I_{[-\infty,\infty]}(z)$
- Kernel Uniform : $K(z)=0,5I_{[-1,1]}(z)$
- Kernel Epanicov : $K(z)=0,75(1-z^{2})I_{[-1,1]}(z)$
- Kernel Kuadrat : $K(z)=\frac{15}{16}(1-z^{2})^{2}I_{[-1,1]}(z)$

dimana $z_{i}=\frac{t-t_{i}}{\alpha},i=1,2,...,n$. t adalah variabel prediktor, t_i adalah nilai ke-i variabel prediktor dan α adalah *bandwidth* . Pendekatan kernel tergantung pada *bandwidth* α, yang berfungsi untuk mengontrol *smoothness* dari kurva estimasi. Pemilihan *bandwitdh* yang tepat merupakan hal yang sangat penting dalam regresi kernel (Hadijati, 2004; Budiantara dan Mulianah, 2007). *Bandwidth* yang terlalu besar akan menghasilkan kurva estimasi yang sangat smooth dan menuju ke rata-rata dari variabel respon, sebaliknya bila *bandwidth* terlalu kecil akan menghasilkan kurva estimasi yang kurang smooth yaitu hasil estimasi akan menuju ke data.

Estimator kurva regresi (2.8) sangat tergantung pada dua hal, yaitu parameter *bandwidth* dan fungsi kernelnya (Budiantara dkk, 2015), tetapi dari dua hal tersebut ternyata *bandwidth* lebih signifikan pengaruhnya terhadap estimator kernel dibanding fungsi kernel. Dalam kaitannya pemilihan *bandwidth* optimal, akan digunakan metode *Generalized Cross Validatian* (GCV). Jika estimasi model (2.11) dinyatakan dengan $\hat{\boldsymbol{g}}_{\alpha}\left(t\right)=\mathbf{D}(\alpha)\mathbf{y}$ maka Fungsi *Generalized Cross Validatian* (GCV) dalam Hong (1999) didefinisikan dengan :

$$GCV(t) = \frac{\frac{1}{n}\left\|(\mathbf{I}-\mathbf{D}(\alpha))\mathbf{y}\right\|^2}{[\frac{1}{n}tr((\mathbf{I}-\mathbf{D}(\alpha))]^2} = \frac{\frac{1}{n}\left\|(\mathbf{I}-\mathbf{D}(\alpha))\mathbf{y}\right\|^2}{[(1-\frac{1}{n}tr\mathbf{D}(\alpha))]^2} \qquad (2.9)$$

Bandwith optimal α diperoleh dengan cara meminimunkan fungsi GCV.

2.5 Estimator Campuran Spline dan Kernel

Budiantara, dkk (2015) meneliti tentang model regresi nonparametrik additif yang memiliki dua komponen variabel prediktor. Komponen prediktor pertama, kurva regresinya dihampiri menggunakan regresi spline, sedang komponen prediktor kedua kurva regresi dihampiri dengan fungsi kernel. Diberikan data berpasangan (x_i, t_i, y_i) data hubungan antar variabel prediktor x_i, t_i dan variabel respon y_i mengikuti model regresi nonparametrik :

$$y_i = \mu(x_i, t_i) + \varepsilon_i \ , \quad \text{i=1,2,…,n}$$

Bentuk kurva regresi diasumsikan tidak diketahui dan hanya diasumsikan smooth dalam arti kontinu dan differensiabel. Error random ε_i berdistribusi normal dengan mean nol dan $E(\varepsilon_i^2) = \sigma^2$. Selanjutnya kurva regesi $\mu(x_i, t_i)$ diasumsikan additif, dalam arti $\mu(x_i, t_i)$ dapat ditulis dalam bentuk :

$$\mu(x_i, t_i) = f(x_i) + g(t_i)$$

Dengan $f(x_i)$ dan $g(t_i)$ merupakan fungsi-fungsi yang smooth. Persoalan utama dalam estimator campuran kurva regresi nonparametrik adalah mendapatkan bentuk estimasi kurva regresi $\mu(x_i, t_i)$ yaitu :

$$\hat{\mu}_{\alpha,\lambda}(x_i, t_i) = \hat{f}_{\alpha,\lambda}(x_i) + \hat{g}_{\alpha}(t_i)$$

Parameter α merupakan parameter bandwith dan λ merupakan titik knot. Estimator campuran regresi spline dan kernel bisa didapatkan dengan cara kurva regresi $f(x_i)$ dihampiri dengan fungsi spline *truncated* derajat m dan titik knot $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_r)^T$. Sedangkan kurva regresi $g(t_i)$ dihampiri dengan fungsi kernel. Menurut Budiantara, dkk (2015), kurva regresi $f(x_i)$ yang diberikan persamaan (2.5) dapat ditulis menjadi :

$$f = X_1\theta + X_2(\lambda)\phi$$

Selanjutnya didapat:

$$\hat{g}_\alpha(t) = D(\alpha)y$$

dengan demikian model regresi campuran spline dan kernel dapat disajikan menjadi :

$$y = [X_1\theta + X_2(\lambda)\phi] + D(\alpha)y + \varepsilon$$

$$= [X_1X_2(\lambda)]\begin{bmatrix}\theta\\ \phi\end{bmatrix} + D(\alpha)y + \varepsilon$$

$$= Z(\lambda)\beta + D(\alpha)y + \varepsilon$$

Estimator parameter β dapat diperoleh dengan menggunakan metode *Ordinary Least Square* (OLS) yang meminimumkan:

$$\sum_{i=1}^{n}\varepsilon_i^2 = Q(\beta/\alpha\,,\lambda)$$

$$= \left\|\left(\mathrm{I} - \mathrm{D}(\alpha)\right)y\right\|^2 - 2\beta^T Z(\lambda)^T\left(\mathrm{I} - \mathrm{D}(\alpha)\right)y + \beta^T Z(\lambda)^T Z(\lambda)\beta$$

Untuk mendapatkan estimator untuk parameter β, dilakukan derivatif parsial terhadap β_j diperoleh :

$$\hat{\beta} = [(Z(\lambda))^T Z(\lambda)]^{-1}(Z(\lambda))^T(I - D(\alpha))y$$

$$\hat{\beta} = C(\lambda, \alpha)y$$

dimana :

$$C(x, \lambda) = \left[\left(Z(\lambda)\right)^T Z(\lambda)\right]^{-1}\left(Z(\lambda)\right)^T\left(I - D(\alpha)\right)$$

Estimator spline diberikan oleh:

$$\hat{f}_{\alpha,\lambda}(x, t) = Z(\lambda)\hat{\beta}(\lambda, \alpha)$$

$$= Z(\lambda)\left[\left(Z(\lambda)\right)^T Z(\lambda)\right]^{-1}\left(Z(\lambda)\right)^T\left(I - D(\alpha)\right)y$$

$$= A(\lambda, \alpha)y\alpha$$

dengan matrik adalah :

$$A(\lambda, \alpha) = Z(\lambda)\left[\left(Z(\lambda)\right)^T Z(\lambda)\right]^{-1}\left(Z(\lambda)\right)^T\left(I - D(\alpha)\right)$$

Selanjutnya didapat estimator campuran regresi spline dan kernel:

$$\hat{\mu}_{\alpha,\lambda}(x, t) = \hat{f}_{\alpha,\lambda}(x, t) + \hat{g}_\alpha(t)$$

$$= A(\lambda, \alpha)y + D(\alpha)y$$

$$= (A(\lambda, \alpha) + D(\alpha))y$$
$$= B(\lambda, \alpha)y$$

dimana matrik B (λ, α)adalah :

$$B(\lambda, \alpha)y = A(\lambda, \alpha) + D(\alpha)$$

Estimator di atas hanya berlaku untuk satu metode spline dan satu kernel. Sedangkan Rismal dkk (2016) memberikan model yang lebih umum, yang berlaku untuk banyak estimator

$$\tilde{\hat{\beta}}\left(\hat{\lambda}, \alpha\right) = \left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1} \left(X\left(\hat{\lambda}\right)\right)^T \left(I - D(\alpha)\right)\tilde{y} \qquad (2.10)$$

Estimator campuran regresi Spline dan Kernel $\hat{\boldsymbol{\mu}}_{\alpha,\lambda}\left(x, t\right)$ sangat bergantung kepada banyak dan letak titik knot $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_r)^T$ dan parameter *bandwitdh* α . Untuk memperoleh estimator campuran regresi Spline dan Kernel yang terbaik perlu dilakukan pemilihan titik knot dan parameter *bandwidth* yang optimal. Metode digunakan adalah *Generalized Cross Validation* (GCV). Fungsi GCV yang diberikan oleh Wahba (1990):

$$GCV\left(\boldsymbol{\lambda}, \alpha\right) = \frac{n^{-1} \| \mathbf{y} - \hat{\boldsymbol{\mu}}_{\alpha,\lambda}(x,t) \|^2}{\left(n^{-1} trace\left[\mathbf{I} - \mathbf{A}\left(\boldsymbol{\lambda}, \alpha\right) - \mathbf{D}(\alpha)\right]\right)^2}$$

Titik knot optimal $\boldsymbol{\lambda}_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)}, ..., \lambda_{r(opt)})^T$ dan parameter *bandwidth* optimal α_{opt} diperoleh dari optimasi :

$$G(\boldsymbol{\lambda}_{opt}, \alpha_{opt}) = \underset{\lambda,\alpha}{Min}\{G(\boldsymbol{\lambda}, \alpha)\}$$

2.6 Koefisien Determinasi ($\mathbf{R^2}$)

Koefisien determinasi ($\mathbf{R^2}$) merupakan ukuran ketelitian atau ketepatan model regresi, atau besarnya kontribusi x terhadap perubahan y. Semakin tinggi nilai R2 akan semakin baik

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

2.7 Interval Konfidensi

Estimasi interval merupakan estimasi pengembangan estimasi titik. Bahwa nilai taksiran parameter tidak berfokus pada satu titik tetapi berdasarkan pada range tertentu, sehingga estimasinya memiliki nilai tertinggi (max) dan nilai terendah (min). interval ini lebih dikenal dengan inteval konfidensi. Nilai yang muncul adalah nilai yang didasarkan probabilitas tertentu, dalam prakteknya biasanya dipilih 90%, 95% atau 99%. Misalkan $X_1, X_2, \dots, X_n$adalah sampel random yang diambil dari populasi dengan parameter θ, maka interval konfidensi $1-\alpha$ untuk θ adalah $P(a \leq \theta \leq b) = 1-\alpha$.

Misalkan interval konfidensi $1-\alpha$ untuk θ adalah $P(a \leq \theta \leq b) = 1-\alpha$ sampel random $X_1, X_2, \dots, X_n$ yang diambil dari populasi yang berdistribusi $N(\mu, \sigma^2)$ dengan σ^2 diketahui. Untuk mendapatkan interval konfidensi $1-\alpha$ untuk parameter μ dapat menggunakan *Pivotal Quantity :*

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Apabila nilai $1-\alpha = 95\%$ maka interval konvidensi 95% untuk parameter μ diberikan oleh :

$$P(-1{,}96 \leq Z \leq 1{,}96) = 95\%$$

$$\Leftrightarrow P\left(-1{,}96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1{,}96\right) = 95\%$$

$$\Leftrightarrow P\left(-1{,}96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

Jika σ tidak diketahui, diganti dengan s yaitu standar deviasi sampel. Dengan demikian, untuk dapat mendapatkan interval konfidensi $1-\alpha$ untuk μdigunakan *Pivotal Quantity :*

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$$

Dengan demikian, interval konfidensi $1-\alpha$untuk parameter μ dapat diperoleh dari :

$$P\left(-t_{(n-1;\, \alpha/2)} \leq T \leq t_{(n-1;\, \alpha/2)}\right) = 1-\alpha$$

$$\Leftrightarrow P\left(-t_{(n-1;\, \alpha/2)} \leq \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \leq t_{(n-1;\, \alpha/2)}\right)$$

$$\Leftrightarrow P\left(\bar{X} - t_{\left(n-1;\, {}^{\alpha}/_{2}\right)}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\left(n-1;\, {}^{\alpha}/_{2}\right)}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Halaman ini sengaja di kosongkan

BAB 3
METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang akan digunakan dalam penelitian ini adalah data sekunder yang berasal dari Badan Pusat Statistika (BPS) tahun 2011. Unit observasi pada penelitian ini adalah 38 Kabupaten/Kota yang ada di Propinsi Jawa Timur pada tahun 2011

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel yang diduga akan berpengaruh terhadap angka kematian bayi yang sesuai dengan penelitian-penelitian sebelumnya. Adapun variabel-variabelnya adalah:

Tabel 3.1 Variabel Penelitian

Variabel	**Nama Variabel**
Y	Angka Kematian Bayi (AKB)
t_1	Persentase Persalinan Yang Dilakukan Dengan Bantuan Non Medis
t_2	Persentase Wanita Perkawinan Pertama Kurang Dari 17 Tahun
t_3	Rata-rata Jumlah Pengeluaran Rumah Tangga Perkapita Sebulan
t_4	Persentase Bayi Usia 0-11 Bulan Diberi Asi
t_5	Jumlah Sarana Kesehatan

Tabel 3.2 Struktur Data

Kabupaten/ Kota	t_1	t_2	t_3	t_4	t_5	$\boldsymbol{y}$
1	t_{11}	t_{21}	t_{31}	t_{41}	t_{51}	y_1
2	t_{12}	t_{22}	t_{32}	t_{42}	t_{52}	y_2
3	t_{13}	t_{23}	t_{33}	t_{43}	t_{53}	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
38	t_{138}	t_{238}	t_{338}	t_{438}	t_{538}	y_{38}

Adapun definisi dari variabel-variabel penelitian diatas menurut definisi istilah yang diterbitkan oleh BPS dan beberapa literatur adalah sebagai berikut :

1. Angka Kematian Bayi (AKB) adalah kematian yang terjadi antara saat setelah bayi lahir sampai bayi belum berumur tepat satu tahun dari setiap 1000 kelahiran hidup pada tahun tertentu (dinyatakan dengan per seribu kelahiran hidup.
2. Persentase pertolongan terakhir persalinan yang dilakukan dengan bantuan non medis adalah jumlah persalinan dengan bantuan non medis dari proses sampai kelahiran bayi di bagi dengan jumlah seluruh persalinan dalam satu tahun dikalikan 100 persen.
3. Persentase perkawinan pada wanita yang usia kawin kurang dari 17 tahun adalah adalah jumlah perkawinan pertama yang usianya kurang dari 17 tahun dibagi dengan jumlah seluruh perkawinan dikalikan dengan 100 persen.
4. Rata-rata jumlah pengeluaran rumah tangga perkapita sebulan adalah rata-rata semua biaya yang dikeluarkan rumah tangga selama sebulan untuk memenuhi kebutuhan konsumsi untuk semua anggota rumah tangga.
5. Persentase bayi usia 0-11 bulan yang di beri ASI adalah jumlah seluruh bayi usia 0-11 bulan yang diberi ASI dibagi dengan jumlah seluruh bayi usia 0-11 bulan dikalikan 100 persen.

6. Jumlah sarana kesehatan (rumah sakit dan puskemas) adalah jumlah seluruh rumah sakit dan puskesmas di setiap kabupaten / kota di propinsi Jawa timur.

3.3 Metode Penelitian

Diberikan data berpasangan $(x_i, t_i, y_i), i = 1,2,3, \dots, n$ yang mengikuti regresi nonparametrik

$$y_i = f(x_i) + g(t_i) + \varepsilon$$

Kurva regresi $f(x_i)$ dihampiri dengan spline truncated:

$$f(x_i) = \sum_{v=0}^{m} \beta_v x_i^v + \sum_{v=1}^{r} \beta_{(v+m)} (x_i - k_v)_+^m$$

1. Mencari interval konfidensi untuk $\beta_v, v = 1,2, \dots, m + r$ pada estimasi campuran regresi nonparametrik spline dan kernel.

   Langkah-langkahnya adalah sebagai berikut:

   1) Menghampiri fungsi $f(x)$ dengan spline truncated dan titik-titik knot $k_1, \dots, k_r$.
   2) Menghampiri fungsi $g(t_i)$ dengan fungsi kernel
   3) Mencari estimasi untuk β, f dan g dalam estimator campuran spline dan kernel.
   4) Mencari distribusi dari $\hat{\beta}$
   5) Dengan menggunakan distribusi dari $\hat{\beta}$ dicari suatu statistik T yang merupakan Pivotal Quantity untuk β
   6) Mencari interval koefisien terpendek untuk parameter β.
2. Memodelkan data angka kematian bayi Propinsi Jawa timur, dengan langkah-langkah sebagai berikut:
   1) Membuat scatter plot antara variabel respon dengan masing-masing variabel prediktor.
   2) Menentukan variabel-variabel prediktor yang menggunakan komponen spline dan komponen kernel.
   3) Memodelkan data menggunakan estimasi campuran spline dan kernel dengan berbagai knot (satu knot, dua knot dan tiga knot).

4) Memilih titik knot dan bandwitdh optimal dengan metode GCV
5) Menetapkan model terbaik dari nilai GCV terkecil.
6) Membuat interval konfidensi terpendek 95% untuk parameter β.
7) Menghitung MSE dan R^2 dari data

BAB IV
HASIL DAN PEMBAHASAN

4.1 Interval Konfidensi Terpendek untuk Parameter Model Komponen Spline Truncated Variansi Diketahui

Diberikan data bepasangan $(x_i, t_i, y_i), i = 1,2,3, \dots, n$ yang mengikuti regresi nonparametrik

$$y_i = f(x_i) + g(t_i) + \varepsilon$$

dengan $f(x_i)$ dan $g(t_i)$ merupakan fungsi yang smooth. Persoalan utama dalam estimator campuran kurva regresi nonparametrik adalah mendapatkan bentuk estimasi kurva regresi y_i.

Untuk mendapatkan estimator campuran regresi Spline dan Kernel, pertama kurva regresi $f(x_i)$ dihampiri dengan fungsi Spline truncated derajat m dan titik knot λ.

$$f(x_i) = \sum_{v=0}^{m} \beta_v x_i^v + \sum_{v=1}^{r} \beta_{(v+m)} (x_i - k_v)_+^m$$

Berdasarkan estimator yang diperoleh pada persamaan 2.10 di peroleh estimasi untuk β, f dan g dalam estimator campuran spline dan kernel berturut-turut adalah

$$\hat{\tilde{\beta}}\left(\hat{\lambda}, \alpha\right) = \left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1} \left(X\left(\hat{\lambda}\right)\right)^T \left(I - D(\alpha)\right) y$$

Dari model diatas di peroleh

$$Y = X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda}, \alpha) + D(\alpha)y + \varepsilon$$

$$Y - D(\alpha)y = X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda}, \alpha) + \varepsilon$$

$$\left(I - D(\alpha)\right)y = X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda}, \alpha) + \varepsilon$$

$$Y = \left(I - D(\alpha)\right)^{-1} \left(X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda}, \alpha) + \varepsilon\right)$$

$$= \left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda}, \alpha) + \left(I - D(\alpha)\right)^{-1} + \varepsilon$$

Karena $Y \sim N\ \left(E(\tilde{Y}),\ Var\ (\tilde{Y})\right)$

$$E(\tilde{Y}) = E\left[\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + \left(I - D(\alpha)\right)^{-1}\varepsilon\right]$$

$$= \left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + 0$$

$$Var\,(\tilde{Y}) = Var\left(\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) + \left(I - D(\alpha)\right)^{-1}\varepsilon\right)$$

$$= 0 + Var[(I - D)^{-1}\varepsilon]$$

$$= \left(I - D(\alpha)\right)^{-1} Var(\varepsilon)\left(I - D(\alpha)\right)^{-1}$$

$$= \left(I - D(\alpha)\right)^{-1}(\sigma^2 I)\left(I - D(\alpha)\right)^{-1}$$

$$= \sigma^2\left(I - D(\alpha)\right)^{-1}\left(I - D(\alpha)\right)^{-1T}$$

$$\tilde{Y} \sim N\left(\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\hat{\tilde{\beta}}(\tilde{\lambda},\alpha), \sigma^2\left(I - D(\alpha)\right)^{-1}\left(I - D(\alpha)\right)^{-1}\right)$$

$$\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) = A\tilde{y},$$

$$A = \left[\left(X(\tilde{\lambda})\right)^T X(\tilde{\lambda})\right]^{-1}\left(X(\tilde{\lambda})\right)^T \left(I - D(\alpha)\right)$$

$$A^T = \left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\,\left[\left(X(\tilde{\lambda})\right)^T X(\tilde{\lambda})\right]^{-1}$$

$$\hat{\tilde{\beta}}(\tilde{\lambda},\alpha) \sim N\left(E\left(\hat{\tilde{\beta}}(\tilde{\lambda},\alpha)\right),\ Var\left(\hat{\tilde{\beta}}(\tilde{\lambda},\alpha)\right)\right)$$

$$E(\hat{\tilde{\beta}}(\tilde{\lambda},\alpha)) = E(A\tilde{y})$$

$$= A\,E(y)$$

$$= \left[\left(X(\tilde{\lambda})\right)^T X(\tilde{\lambda})\right]^{-1}\left(X(\tilde{\lambda})\right)^T \left(I - D(\alpha)\right)E(\tilde{y})$$

$$= \left[\left(X(\tilde{\lambda})\right)^T X(\tilde{\lambda})\right]^{-1}\left(X(\tilde{\lambda})\right)^T \left(I - D(\alpha)\right)\left(I - D(\alpha)\right)^{-1} X(\tilde{\lambda})\beta$$

$$= \beta$$

$$Var\left(\hat{\tilde{\beta}}\right) = Var\,(Ay) = A\,Var\,(y)A^T$$

$$Var\left(\hat{\tilde{\beta}}\right) = Var\left(\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}\left(X\left(\hat{\lambda}\right)\right)^T \left(I - D(\alpha)\right)\tilde{y}\right)$$

$$= \left(\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}\left(X\left(\hat{\lambda}\right)\right)^T \left(I - D(\alpha)\right)\right) Var(\tilde{y})$$

$$\left(I - D(\alpha)\right)^T X\left(\hat{\lambda}\right)\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

$$Var\left(\hat{\tilde{\beta}}\right)=\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}\left(X\left(\hat{\lambda}\right)\right)^T\left(I-D(\alpha)\right)\sigma^2\left(I-D(\alpha)\right)^{-1}$$

$$\left(I-D(\alpha)\right)^{-1T}\left(I-D(\alpha)\right)^T X\left(\hat{\lambda}\right)\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

$$=\sigma^2\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}\left(X\left(\hat{\lambda}\right)\right)^T\left(I-D(\alpha)\right)\sigma^2\left(I-D(\alpha)\right)^{-1}$$

$$\left(I-D(\alpha)\right)^{-1T}\left(I-D(\alpha)\right)^T X\left(\hat{\lambda}\right)\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

$$=\sigma^2\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

$$=\sigma^2 W\left(\hat{\lambda},\alpha\right)$$

$$W=\left[\left(X\left(\hat{\lambda}\right)\right)^T X\left(\hat{\lambda}\right)\right]^{-1}$$

$$\left(\hat{\tilde{\beta}}\left(\hat{\lambda},\alpha\right)\right)\sim N\left[E\left(\hat{\tilde{\beta}}\left(\hat{\lambda},\alpha\right)\right), Var\hat{\tilde{\beta}}\left(\hat{\lambda},\alpha\right)\right]$$

$$\left(\hat{\tilde{\beta}}\left(\hat{\lambda},\alpha\right)\right)\sim N\left[\beta,\sigma^2 W\left(\hat{\lambda},\alpha\right)\right]$$

Langkah-langkah interval konfidensi

$$Z=\frac{\hat{\tilde{\beta}}_j-E\left(\hat{\tilde{\beta}}_j\right)}{\sqrt{Var\left(\hat{\tilde{\beta}}_j\right)}}\sim N\left(0,1\right)$$

$$=\frac{[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1}X(\tilde{\lambda})^T(I-D(\alpha))\tilde{y}]_j-\beta_j}{[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj}}\sim N\left(0,1\right)$$

dengan ω_j merupakan elemen diagonal ke- j dari suatu matriks $W(\tilde{\lambda},\alpha)$. variabel random Z_j berdistribusi $N(0,1)$. Dengan demikian, Z_j merupakan pivotal Quantity untuk kurva regresi β_j. Interval konfidensi $1-\alpha$ diperoleh dari menyelesaikan persamaan probabilitas:

$$P\left(a\leq Z_j\leq b\right)=1-\alpha$$

dengan $a\,\epsilon\,R,\ b\,\epsilon\,R\ \ a<b,\ i=1,2,\dots,n$. Persamaan diatas dapat dinyatakan menjadi:

$$P\left(a\leq\frac{[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1}X(\tilde{\lambda})^T(I-D(\alpha))\tilde{y}]_j}{[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj}}\leq b\right)=1-\alpha$$

Dengan sedikit penjabaran, sehingga diperoleh interval konfidensi $1-\alpha$ untuk $\beta_j, j = 1,2,\ldots,n$

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j + b[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j + a[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1-\alpha$$

Dengan konsep interval terpendek harus ditentukan nilai $a\epsilon R\ dan\ b\epsilon R$, sehingga panjang interval $\lambda(a,b)$ pada persamaan diatas terpendek. Untuk tujuan ini dicari penyelesaian optimasi bersyarat berikut.

$$\min_{a\epsilon R\ ,b\epsilon R}\{\lambda(a,b)\} = \min_{a\epsilon R\ ,b\epsilon R}\left\{(b-a) \quad \left[\sqrt{\sigma^2\omega_i(\tilde{\lambda},\alpha)}\right]_{jj}\right\} \tag{4.2}$$

Dengan syarat $\int_a^b \varphi(z)dz = 1-\alpha$, atau $\phi(b)-\phi(a)-(1-\alpha) = 0$ (4.3)

Fungsi φ merupakan distribusi probabilitas N(0,1) dan Φ merupakan distribusi probabilitas kumulatif N(0,1). Optimasi 1 dan 2 dapat diselesaikan dengan menggunakan metode lagrange multiple dibentuk fungsi lagrange multiple.

$$\Omega(a,b,c) = (b-a)\left[\sqrt{\sigma^2\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} + c\left[\int_a^b \varphi(z)dz - (1-\alpha)\right]$$

Selanjutnya dengan menderivatifkan fungsi $\Omega(a,b,c)$terhadap a, b dan c diperoleh:

$$\frac{\Omega(a,b,c)}{\partial a} = 0 \Longrightarrow -\left[\sqrt{\sigma^2\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} - c\varphi(a) = 0 \tag{4.4}$$

$$\frac{\Omega(a,b,c)}{\partial b} = 0 \Longrightarrow \left[\sqrt{\sigma^2\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} + c\ \varphi(b) = 0 \tag{4.5}$$

$$\frac{\Omega(a,b,c)}{\partial c} = 0 \Longrightarrow \Omega(b)-\Omega(a)-(1-\alpha) = 0 \tag{4.6}$$

Persamaan 4.4 dan 4.5 menghasilkan penyelesaian

$$\varphi(a) = \varphi(b) \tag{4.7}$$

Menginggat persamaan 4.6 dan $Z\sim N(0,1)$ maka penyelesaian persamaan 4.7 adalah $a = b$ atau $a = -b$. Akan tetapi, persamaan $a = b$ tidak memenuhi. Jadi,agar diperoleh interval konfidensi terpendek harus harus diambil nilai a dan b yang memenuhi persamaan

$$\int_{-\infty}^{a} \varphi(z)dz = \frac{\alpha}{2} = \int_{b}^{\infty} \varphi(z)dz \tag{4.8}$$

Jika tingkat konfidensi $1-\alpha$ diberikan maka nilai a dan b dapat dilihat pada tabel distribusi $N(0,1)$. Interval konfidensi $1-\alpha$ untuk kurva regresi β_j, $j = 1,2,\dots,n$ diberikan oleh persamaan 4.1 dengan a dan b memenuhi persamaan (4.8)

4.2 Interval Konfidensi Terpendek untuk Parameter Model Komponen Spline Truncated Variansi Tidak Diketahui

Umumnya, dalam analisis regresi spline, varians σ^2 tidak diketahui. Jika demikian σ^2 dapat diganti dengan estimasinya, yaitu:

$$\hat{\sigma}^2 = MSE = \frac{1}{n-m-r-1}\sum_{i=1}^{n}\left(\hat{y}-y_i\right)^2$$

dimana :

$$\begin{aligned}
\sum_{i=1}^{n}\left(\hat{y}-y_i\right)^2 &= y^T y - y^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1-D(\alpha)\right)y - y^T D(\alpha)y - y^T \\
&\left(1-D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T y + y^T\left(1-D(\alpha)\right)^T X(\lambda) \\
&\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T\left(1-D(\alpha)\right)y \\
&+ y^T\left(1-D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T Dy - y^T D^T y \\
&+ y^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T\left(1-D(\alpha)\right)y + D^T y^T Dy \\
&= y^T (I - X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T\left(1-D(\alpha)\right) - D(\alpha) - \\
&\left(1-D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T + \left(1-D(\alpha)\right)^T X(\lambda) \\
&\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T\left(1-D(\alpha)\right)y \\
&+ \left(1-D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T D - D^T \\
&+ X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T\left(1-D(\alpha)\right) + D^T D)\,\mathrm{y}
\end{aligned}$$

$$= y^T(I - X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1 - D(\alpha)\right) - D(\alpha) -$$

$$\left(1 - D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T + \left(1 - D(\alpha)\right)^T X(\lambda)$$

$$\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1 - D(\alpha)\right) y$$

$$+ \left(1 - D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T D - D^T$$

$$+ X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1 - D(\alpha)\right) + D^T D)y$$

$$= y^T A y$$

$$A = (I - X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1 - D(\alpha)\right) - D(\alpha) -$$

$$\left(1 - D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T + \left(1 - D(\alpha)\right)^T X(\lambda)$$

$$\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1 - D(\alpha)\right) y$$

$$+ \left(1 - D(\alpha)\right)^T X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1T} X(\lambda)^T D - D^T$$

$$+ X(\lambda)\left(X(\lambda)^T X(\lambda)\right)^{-1} X(\lambda)^T \left(1 - D(\alpha)\right) + D^T$$

$$T = \frac{\hat{\tilde{\beta}}_j - E\left(\hat{\tilde{\beta}}\right)}{\sqrt{Var\left(\hat{\tilde{\beta}}_j\right)}} \sim t_{(n-m-r-1)}$$

$$T = \frac{\hat{\tilde{\beta}}_j - \beta_j}{\sqrt{Var\left(\hat{\tilde{\beta}}_j\right)}}$$

$$\hat{\tilde{\beta}}_j \sim N\left(\beta, \sigma^2 \omega_j\left(\tilde{\lambda}, \alpha\right)\right)$$

$$T = \frac{\left[\left[X\left(\tilde{\lambda}\right)^T X\left(\tilde{\lambda}\right)\right]^{-1} X\left(\tilde{\lambda}\right)^T \left(I - D(\alpha)\right)\tilde{y}\right]_j - \beta_j}{\left[\sqrt{MSE\,\omega_j\left(\tilde{\lambda}, \alpha\right)}\right]_{jj}}$$

$$T = \frac{\dfrac{\left[\left[X\left(\tilde{\lambda}\right)^T X\left(\tilde{\lambda}\right)\right]^{-1} X\left(\tilde{\lambda}\right)^T \left(I - D(\alpha)\right)\tilde{y}\right]_j - \beta_j}{\sqrt{\sigma^2 \omega_j\left(\tilde{\lambda}, \alpha\right)}}}{\left[\dfrac{MSE}{\sigma^2}\right]}$$

$$T=\frac{\left[\left[X\left(\tilde{\lambda}\right)^{T}X\left(\tilde{\lambda}\right)\right]^{-1}X\left(\tilde{\lambda}\right)^{T}\left(I-D(\alpha)\right)\tilde{y}\right]_{j}-\beta_{j}}{\left[\sqrt{MSE\omega_{j}\left(\tilde{\lambda},\alpha\right)}\right]_{jj}}$$

$$T=\frac{\dfrac{\left[\left[X\left(\tilde{\lambda}\right)^{T}X\left(\tilde{\lambda}\right)\right]^{-1}X\left(\tilde{\lambda}\right)^{T}\left(I-D(\alpha)\right)\tilde{y}\right]_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}\left(\tilde{\lambda},\alpha\right)}}}{\left[\dfrac{MSE}{\sigma^{2}}\right]}$$

$$T=\frac{\dfrac{\left[\left[X\left(\tilde{\lambda}\right)^{T}X\left(\tilde{\lambda}\right)\right]^{-1}X\left(\tilde{\lambda}\right)^{T}\left(I-D(\alpha)\right)\tilde{y}\right]_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}\left(\tilde{\lambda},\alpha\right)}}}{\left[\dfrac{\dfrac{\sum_{i=1}^{n}\left(\tilde{y}_{i}-y_{i}\right)}{\sigma^{2}}}{n-m-r-1}\right]_{jj}}$$

Misalkan $A=\frac{\left[\left[\left(X(\tilde{\lambda})\right)^{T}X(\tilde{\lambda})\right]^{-1}\left(X(\tilde{\lambda})\right)^{T}(I-D(\alpha))\tilde{y}\right]_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}(\tilde{\lambda},\alpha)}}$ dan $B=\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{i}-y_{i})^{2}}{\sigma^{2}}}$

$$A=\frac{\left[\left[X\left(\tilde{\lambda}\right)^{T}X\left(\tilde{\lambda}\right)\right]^{-1}X\left(\tilde{\lambda}\right)^{T}\left(I-D(\alpha)\right)\tilde{y}\right]_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}\left(\tilde{\lambda},\alpha\right)}}$$

$$E(A)=\frac{\left[\left[X\left(\tilde{\lambda}\right)^{T}X\left(\tilde{\lambda}\right)\right]^{-1}X\left(\tilde{\lambda}\right)^{T}\left(I-D(\alpha)\right)\tilde{y}\right]_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}\left(\tilde{\lambda},\alpha\right)}}$$

$$=\frac{E\left[\left[X\left(\tilde{\lambda}\right)^{T}X\left(\tilde{\lambda}\right)\right]^{-1}X\left(\tilde{\lambda}\right)^{T}\left(I-D(\alpha)\right)\tilde{y}\right]_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}\left(\tilde{\lambda},\alpha\right)}}$$

$$=\frac{\hat{\tilde{\beta}}_{j}-\beta_{j}}{\sqrt{\sigma^{2}\omega_{j}\left(\tilde{\lambda},\alpha\right)}}=0$$

$$Var(A) = Var\left(\frac{\left[\left[X(\tilde{\lambda})^T X(\tilde{\lambda}) \right]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y} \right]_j - \beta_j}{\sqrt{\sigma^2 \omega_j(\tilde{\lambda}, \alpha)}} \right)$$

$$= \left(\frac{1}{\sqrt{\sigma^2 \omega_j(\tilde{\lambda}, \alpha)}} \right)^2 Var\left(\left[\left[X(\tilde{\lambda})^T X(\tilde{\lambda}) \right]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y} \right]_j - \beta_j \right)$$

$$= \frac{1}{\sigma^2 \omega_j(\tilde{\lambda}, \alpha)} Var\left(\left[\left[X(\tilde{\lambda})^T X(\tilde{\lambda}) \right]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y} \right]_j \right)$$

$$= \frac{1}{\sigma^2 \omega_j(\tilde{\lambda}, \alpha)} \left(\sigma^2 \omega_j(\tilde{\lambda}, \alpha) \right)$$

$$= 1$$

Dari hasil penjabaran diatas di peroleh $A \sim N(0,1)$

$$B = \frac{\sum_{i=1}^{n} (\tilde{y}_i - y_i)^2}{\sigma^2}$$

$$= \frac{y^T A y}{\sigma^2} \sim \chi^2_{(n-m-r-1)}$$

$$B \sim \chi^2_{(n-m-r-1)}$$

Karena $A \sim N(0,1)$ dan $B \sim \chi^2_{n-m-r-1}$ maka`` dapat disimpulkan bahwa A dan B adalah independen.

$$T = \frac{[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j - \beta_j}{[\sqrt{MSE\omega_j(\tilde{\lambda}, \alpha)}]_{jj}} \sim t_{(n-m-r-1)}$$

Dengan ω_j elemen diagonal ke- j dari matriks $W(\tilde{\lambda}, \alpha)$. Variabel random T berdistribusi $t_{(n-m-r-1)}$. Dengan demikian, T merupakan pivotal Quantity untuk kurva regresi β_j. Interval konfidensi $1 - \alpha$ diperoleh dari menyelesaikan persamaan probabilitas:

$$P(a^* \leq T \leq b^*) = 1 - \alpha$$

$a^* \epsilon R,\ b \epsilon R \quad a < b^*,\ j = 1,2, \dots, m + r$

persamaan diatas dapat dinyatakan menjadi:

$$P\left(a^* \le \frac{[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j - \beta_j}{[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}} \le b^* \right) = 1 - \alpha$$

Setelah dilakukan penjabaran, maka diperoleh interval konfidensi $1 - \alpha$ untuk $\beta_j, j = 1,2, \dots, m + r$

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + b^*[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + a^*[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1 - \alpha$$

Dengan konsep interval terpendek harus ditentukan nilai $a^* \epsilon R \ dan \ b^* \epsilon R$ sehingga panjang interval $\lambda(a^*, b^*)$ pada persamaan diatas terpendek. Untuk tujuan ini dicari penyelesaian optimasi bersyarat berikut.

$$\min_{a\epsilon R ,b\epsilon R}\{\lambda(a^*, b^*)\} = \min_{a\epsilon R ,b\epsilon R}\left\{(b^* - a^*) \quad \left[\sqrt{MSE\omega_i(\tilde{\lambda},\alpha)}\right]_{jj}\right\} \qquad (4.10)$$

Dengan syarat

$$\int_{a^*}^{b^*} \varphi(z)dz = 1 - \alpha, \text{ atau } \phi(b^*) - \phi(a^*) - (1 - \alpha) = 0 \qquad (4.11)$$

Fungsi φ merupakan distribusi probabilitas $t_{(n-m-r-1)}$dan φ merupakan distribusi probabilitas kumulatif $t_{(n-m-r-1)}$.Optimasi 1 dan 2 dapat diselesaikan dengan menggunakan metode lagrange multiple dibentuk fungsi lagrange multiple

$$\Omega(a^*, b^*, c) = (b^* - a^*)\left[\sqrt{MSE\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} + c\left[\int_{a^*}^{b^*} \varphi(z)dz - (1 - \alpha)\right]$$

Selanjutnya dengan menderivatifkan fungsi $\Omega(a^*, b^*, c)$terhadap a^*, b^* dan c diperoleh:

$$\frac{\Omega(a^*,b^*,c)}{\partial a^*} = 0 \Longrightarrow -\left[\sqrt{MSE\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} - c\varphi(a^*) = 0 \qquad (4.12)$$

$$\frac{\Omega(a^*,b^*,c)}{\partial b^*} = 0 \Longrightarrow \left[\sqrt{MSE\omega_i(\tilde{\lambda},\alpha)}\right]_{jj} + c\ \varphi(b^*) = 0 \qquad (4.13)$$

$$\frac{\Omega(a^*,b^*,c)}{\partial c} = 0 \Longrightarrow \Omega(b^*) - \Omega(a^*) - (1 - \alpha) = 0 \qquad (4.14)$$

Persamaan 4.4 dan 4.5 menghasilkan penyelesaian

$$\varphi(a^*) = \varphi(b^*) \qquad (4.15)$$

Menginggat persamaan (4.14) dan $T \sim t_{(n-m-r-1)}$ maka penyelesaian persamaan (4.15) adalah $a^* = b^*$ atau $a^* = -b^*$. Akan tetapi, persamaan $a^* = b^*$ tidak memenuhi. Jadi,agar diperoleh interval konfidensi terpendek harus harus diambil nilai a dan b yang memenuhi persamaan

$$\int_{-\infty}^{a^*} \varphi(t)dt = \frac{\alpha}{2} = \int_{b^*}^{\infty} \varphi(t)dt \tag{4.16}$$

Jika tingkat konfidensi $1 - \alpha$ diberikan maka nilai a dan b dapat dilihat pada tabel distribusi $t_{(n-m-r-1)}$. Interval konfidensi $1 - \alpha$ untuk kurva regresi β_j, $j = 1,2, \ldots, n$ diberikan oleh persamaan (4.9) dengan a^* dan b^* memenuhi persamaan 4.16.

4.3. Aplikasi Pada Model Angka Kematian Bayi Di Provinsi Jawa Timur

4.3.1. Analisis Deskriptif Angka Kematian Bayi Propinsi Jawa Timur

Angka Kematian Bayi di propinsi Jawa Timur masih menunjukan angka yang tinggi untuk Indonesia. Selain itu dengan karakteristik yang beragam untuk wilayah kabupaten kota di propinsi jawa timur sehingga mengakibatkan beragam pula untuk AKB nya. Pada penelitian ini menggunakan variabel respon (y) angka kematian bayi dengan 5 variabel prediktor yang diduga mempengaruhi Angka kematian bayi di Jawa Timur. Analisis pertama yang dipergunakan adalah analisis deskriptif untuk mengetahui karakteristik AKB dengan ke lima variabel prediktor yang diduga mempengaruhinya. Berdasarkan Tabel 4.1 diketahui rata-rata AKB di propinsi Jawa timur adalah 34,18 artinya terjadi kematian bayi sebesar 34 setiap terjadi 1.000 kelahiran bayi. AKB maksimal sebesar 64,19 dan minimum adalah 20,02 dengan keragaman sebesar 160,7.

Dalam penelitian ini memfokuskan untuk menyelidiki pola hubungan perilaku data AKB (y) terhadap beberapa variabel yang dianggap berpengaruh. Beberapa variabel yang dianggap berpengaruh diantaranya yaitu variabel persentase persalinan yang dilakukan dengan bantuan non medis (x_1), untuk persentase wanita perkawinan pertama kurang dari 17 tahun (x_2) , untuk rata-rata jumlah pengeluaran rumah tangga perkapita sebulan (x_3), Variabel yang ke empat adalah persentase bayi usia 0-11 bulan di beri ASI (x_4), dan jumlah sarana kesehatan (x_5). Selanjutnya guna mendapatkan estimasi model yang sesuai dengan pola data dari variabel respon terhadap masing-masing variabel prediktornya, digunakan langkah–langkah yang telah dipaparkan pada bab 3 sebelumnya. Adapun analisis pertama yang dipergunakan adalah analisis deskriptif untuk mengetahui karakteristik AKB dengan ke lima variabel prediktor yang diduga mempengaruhinya dimasing-masing kabupaten. Secara lengkap statistika deskriptif dari masing-masing kabupaten ditunjukkan pada tabel dibawah :

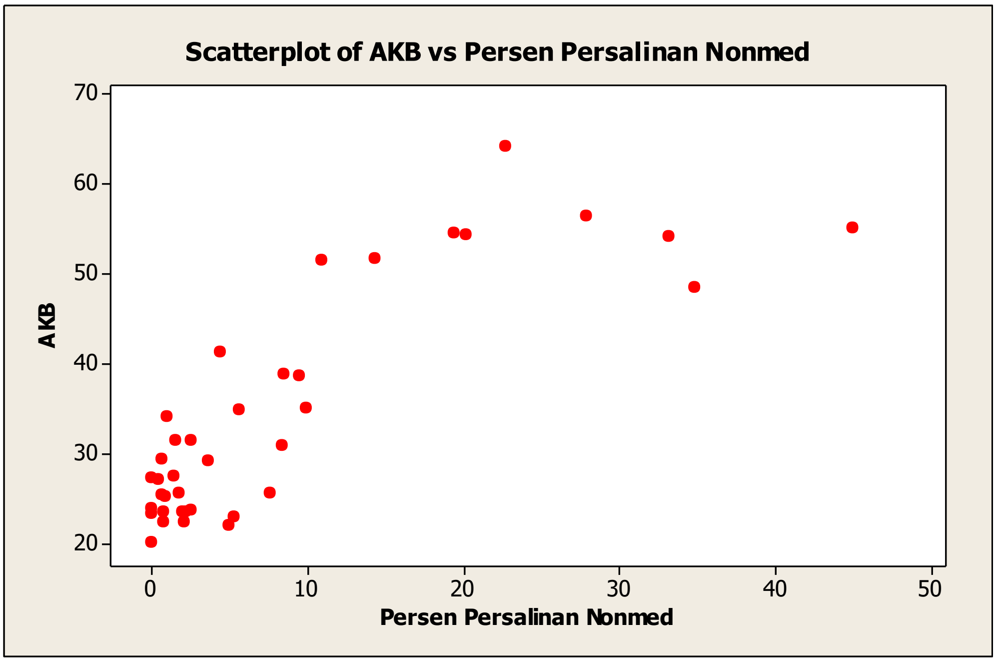
Tabel 4.1 Statistika Deskriptif AKB dengan 5 Variabel Prediktor

Variabel	**Mean**	**Maksimum**	**Minimum**	**Varians**
$\boldsymbol{Y}$	34,18	64,19	20,02	160,7
$\boldsymbol{x_1}$	8,31	44,99	0,00	125,29
$\boldsymbol{x_2}$	27,01	59,09	10,07	169,11
$\boldsymbol{x_3}$	239,3	558.6	122,2	11164,1
$\boldsymbol{x_4}$	93,95	100,00	83,92	21,532
$\boldsymbol{x_5}$	0,10753	0,00138	0,04866	0,19856

Lebih lanjut dari Tabel 4.1 memperlihatkan bahwa masing-masing kabupaten memiliki karakteristik yang berbeda-beda untuk semua variabel prediktornya. Kabupaten Probolinggo merupakan daerah dengan variabel persentase persalinan yang dilakukan dengan bantuan non medis (x_1) diperoleh rata-rata 8,31% , untuk persentase wanita perkawinan pertama kurang dari 17 tahun (x_2) diperoleh rata rata 27,01%, untuk rata-rata jumlah pengeluaran rumah tangga perkapita sebulan (x_3) adalah Rp.239.322,00 dengan keragaman yang cukup besar. Variabel yang ke empat adalah persentase bayi usia 0-11 bulan di beri ASI (x_4), dengan rata-rata sebesar 93,95% dan jumlah sarana kesehatan (x_5) dengan rata-rata 90,10.

4.3.2. Model AKB dengan Propinsi Jawa Timur

Angka Kematian Bayi propinsi Jawa Timur dengan variabel prediktornya akan di jelaskan pola hubungannya pada sub bab ini. Pola hubungan 34 ini menjadi dasar untuk membentuk model regresinya. Pola hubungan AKB dengan variabel prediktor pertama yaitu persentase persalinan dengan bantuan tenaga nonmedis diberikan pada Gambar (4.1).

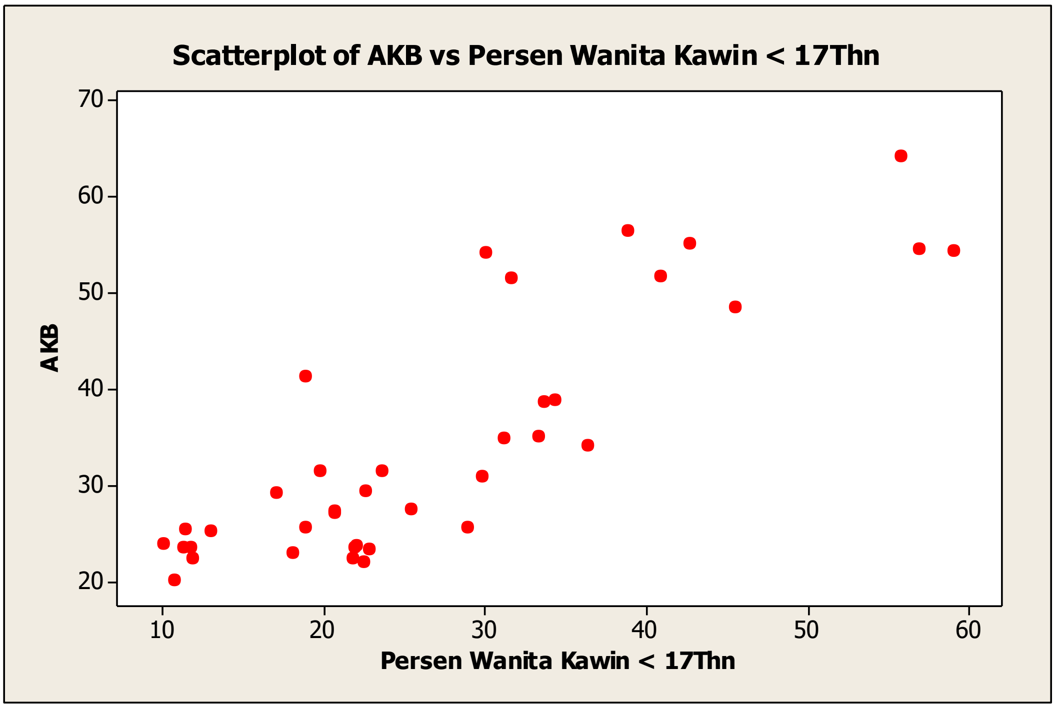


Gambar 4.1 *Plot* antara AKB dengan persentase persalinan dengan bantuan nonmedis

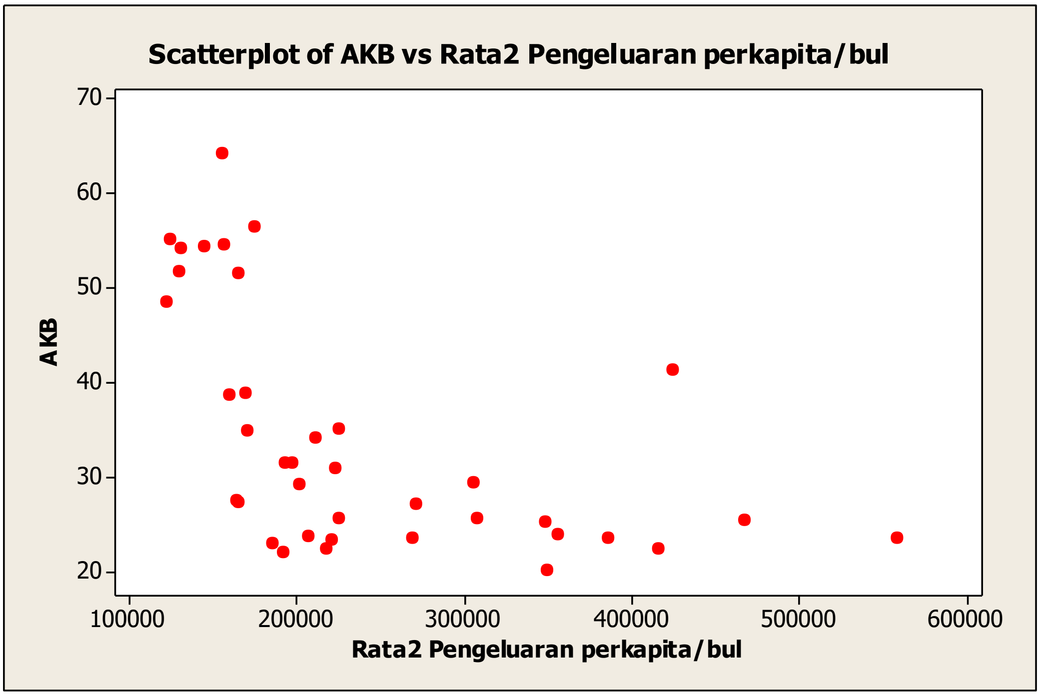
Dari gambar 4.1 tampak pola hubungan antara AKB dan variabel persentase persalinan ditolong dengan tenaga non medis tidak ada kecenderungan membentuk pola tertentu. AKB tertinggi tidak diperoleh dari persentase yang tinggi atau yang rendah. . Pola hubungan yang berubah-ubah pada sub-sub interval tertentu dapat didekati dengan regresi spline *truncated* .

Selanjutnya akan dilihat pola hubungan AKB dengan variabel prediktor yang kedua yaitu persentase perkawinan pertama wanita kurang dari 17 tahun.

Pada Gambar 4.2 terlihat bahwa AKB dan persentase wanita kawin < 17 tahun tidak membentuk kecenderungan pola tertentu. AKB tertinggi ternyata tidak diperoleh dari persentase tertinggi wanita kawin <17 tahun dan sebaliknya AKB terendah tidak diperoleh dari persentase paling kecil dari wanita kawin < 17 tahun. Pola hubungan yang berubah-ubah pada sub-sub interval tertentu dapat didekati dengan regresi spline *truncated*.



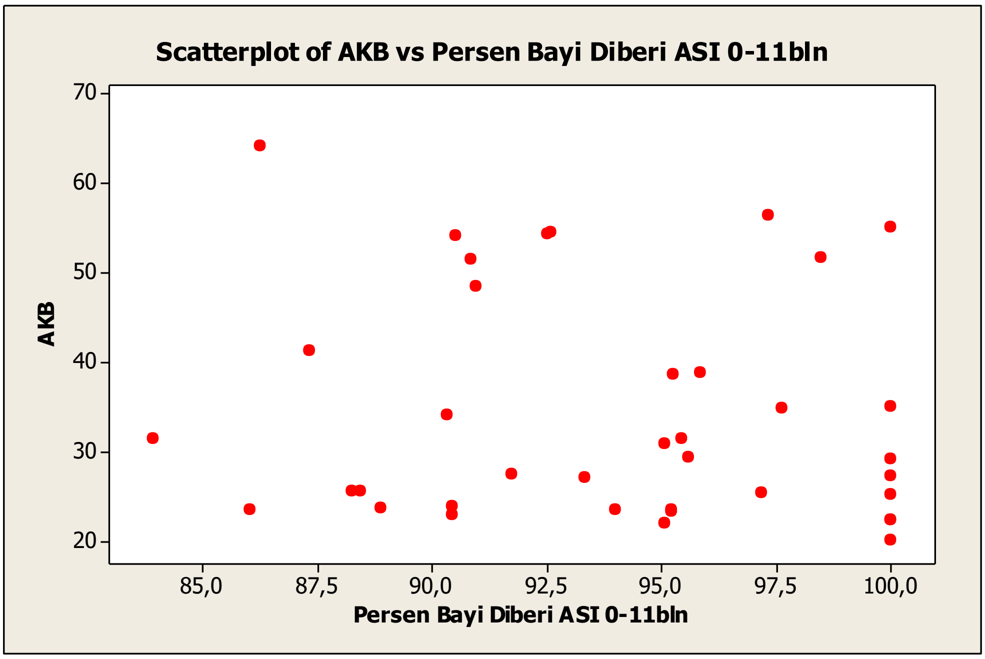
Gambar 4.2 *Plot* Antara AKB dengan Persentase wanita perkawinan pertama < 17 Tahun



Gambar 4.3 *Plot* AKB dengan rata-rata jumlah pengeluaran rumah tangga perkapita sebulan

Pola hubungan AKB dengan variabel prediktor ke tiga yaitu rata-rata jumlah pengeluaran rumah tangga perkapita sebulan dapat dilihat pada Gambar 4.3. Pada *plot* antara AKB dan rata-rata jumlah pengeluaran rumah tangga perkapita sebulan terlihat bahwa pengeluaran perkapita sebulan rendah akan menyebabkan nilai AKB yang tinggi dan juga rendah dan rata-rata pengeluaran perkapita sebulan tinggi dan sedang menyebabkan AKB yang rendah juga. Sehingga pola hubungan kedua variabel tersebut cenderung tidak memiliki pola hubungan tertentu. Pola hubungan yang berubah-ubah pada sub-sub interval tertentu dapat didekati dengan regresi spline *truncated*.

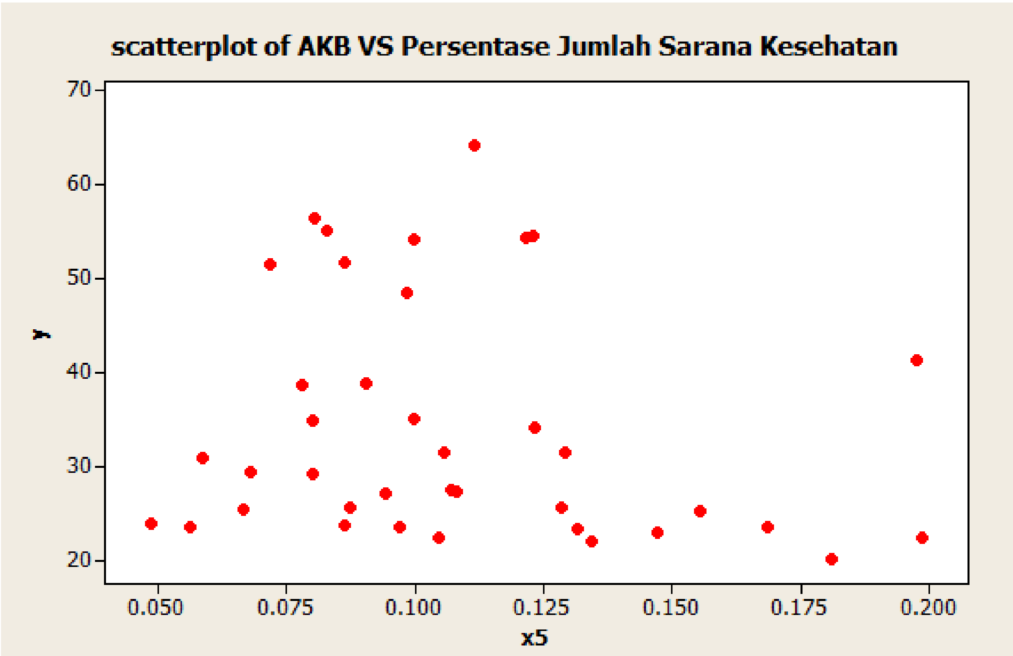
Selanjutnya akan dilihat pola hubungan AKB dengan variabel prediktor yang keempat yaitu Persentase bayi usia 0-11 bulan di beri ASI. Pola hubungan dapat dilihat pada Gambar 4.4. Pada *plot* dapat terlihat bahwa persentase bayi usia 0-11 bulan diberi ASI akan cenderung menghasilkan nilai AKB yang tinggi dan juga rendah. Persentase bayi usia 0-11 bulan diberi ASI tinggi juga akan menghasilkan nilai AKB yang tinggi dan juga rendah. Sehingga pola hubungan antara AKB dan persentase bayi usia 0-11 bulan diberi ASI cenderung tidak berpola tertentu sehingga Pola hubungan yang tidak jelas dapat didekati dengan regresi kernel.



Gambar 4.4 *Plot* AKB dengan persentase bayi usia 0-11 bulan di beri ASI

Selanjutnya akan dilihat pola hubungan AKB dengan variabel prediktor yang kelima yaitu Jumlah sarana kesehatan. Pola hubungan ini dapat dilihat pada Gambar 4.5. Jumlah sarana kesehatan yang sedang cenderung menghasilkan nilai

AKB yang tinggi dan juga rendah. Nilai AKB yang rendah juga cenderung dihasilkan dari jumlah sarana kesehatan yang tinggi dan juga rendah. Pada *plot* pencaran data pola hubungan antara AKB dan jumlah sarana kesehatan cenderung tidak memiliki pola hubungan tertentu, sehingga Pola hubungan yang tidak jelas dapat didekati dengan regresi kernel.



Gambar 4.5 *Plot* AKB dengan Jumlah Sarana Kesehatan

Pada *scatterplot* yang diperoleh dari lima variabel prediktor terhadap variabel respon terlihat bahwa pola data menyebar dan tidak membentuk pola terentu. Pada *scatter plot* juga terlihat ada indikasi ketaksamaan variansi (kasus heteroskedastik).

4.3.3. Pemilihan Titik Knot Optimal Pada Spline Linier

Langkah awal yang harus dilakukan adalah menentukan lokasi dari titik *knot*. Cara pemilihan *knot* dengan cara mencoba-coba pada tempat yang terlihat adanya perubahan perilaku data pada selang tertentu. Dari berbagai *knot* dipilih nilai GCV minimum. Nilai GCV minimum adalah nilai yang dipergunakan untuk menentukan titik knot optimal. Untuk mendapatkan hasil yang terbaik dalam regresi *nonparametrik* dicoba model campuran *spline truncated* dan *kenel* dengan satu *knot* sampai tiga *knot*. Selanjutnya menghitung nilai estimasi parameter dan membuat kurva estimasinya. Untuk perhitungan GCV dilakukan dengan fasilitas program R. Akan dicari GCV untuk 1 knot, 2 knot dan 3 knot pada model campuran spline truncated dan kernel. Adapun modelnya dapat kita lihat sebagai berikut

4.3.3.1.Pemilihan Titik Knot Optimal *Spline Truncated* dan kernel 1 Titik Knot

Pada bagian ini dibahas pemilihan titik knot optimal pada regresi spline truncated 1 titik knot pada AKB Propinsi Jawa Timur dengan lima variabel prediktornya. Ringkasan perhitungan pemilihan titik knot dan nilai GCV diberikan pada Tabel 4.2.

Tabel 4.2 Skor GCV untuk 1 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Banyak knot	Lokasi knot			Bandwith		GCV	R^2
	x_1	x_2	x_3	α_1	α_2		
	k_1	k_1	k_1				
Knot 1	10.100	21.074	220196.1	3.620	39.520	33.678	84,98%
	11.018	22.075	229101.2	3.948	43.112	33.873	
	9.182	20.074	211291.0	3.292	35.928	33.915	
	11.936	23.075	238006.3	4.276	46.704	34.011	
	12.854	24.076	246911.4	4.604	50.296	34.348	

Ringkasan perhitungan pemilihan titik knot dan nilai GCV diberikan pada Tabel 4.3. Berdasarkan Tabel 4.3 dapat diketahui nilai GCV minimum adalah 31,92142. Nilai GCV minimum ini memberikan titik knot:

x_1 = 10.100 (untuk variabel persentase persalinan yang dilakukan dengan bantuan non medis)

x_2= 21.074 (untuk variabel persentase wanita perkawinan pertama kurang dari 17 tahun)

x_3 = 220196.100 (untuk variabel rata-rata jumlah pengeluaran rumah tangga perkapita sebulan)

Pemilihan model campuran regresi nonparametrik spline truncated dan kernel terbaik ditentukan berdasarkan titik-titik knot yang optimum yang didapatkan dari nilai GCV yang paling kecil. Berikut ini adalah estimasi model campuran regresi nonparametrik spline truncated dan kernel dengan 1 titi knot :

$$\hat{y}=\hat{\beta}_{10}+\hat{\beta}_{11}x_{1i}+\hat{\beta}_{12}\left(x_{1i}-k_{1i}\right)_{+}+\hat{\beta}_{20}+\hat{\beta}_{21}x_{2i}+\hat{\beta}_{22}\left(x_{2i}-k_{2i}\right)_{+}-\hat{\beta}_{30}+\hat{\beta}_{31}x_{3i}$$
$$+\hat{\beta}_{32}\left(x_{3i}-k_{3i}\right)_{+}+\sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_1}K\left(\frac{t_1-t_{1i}}{\alpha_1}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_1}K\left(\frac{t_1-t_{1i}}{\alpha_1}\right)}\right)y_i+\sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_2}K\left(\frac{t_2-t_{2i}}{\alpha_2}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_2}K\left(\frac{t_2-t_{2i}}{\alpha_2}\right)}\right)y_i$$

$$\hat{y}=0,0156+0,3569x_{1i}+0,0156\left(x_{1i}-10,100\right)_{+}+0,6209+0,0156x_{2i}$$
$$-0,0001\left(x_{2i}-21,074\right)_{+}-0,0522-0,1227x_{3i}+0,0001\left(x_{3i}-220196,100\right)_{+}$$
$$+\sum_{i=1}^{38}\left(\frac{\frac{1}{3,6197}K\left(\frac{t_1-t_{1i}}{3,6197}\right)}{\sum_{j=1}^{38}\frac{1}{3,6197}K\left(\frac{t_1-t_{1i}}{3,6197}\right)}\right)y_i+\sum_{i=1}^{R}\left(\frac{\frac{1}{39,5202}K\left(\frac{t_2-t_{2i}}{39,5202}\right)}{\sum_{j=1}^{R}\frac{1}{39,5202}K\left(\frac{t_2-t_{2i}}{39,5202}\right)}\right)y_i$$

4.3.3.2.Pemilihan Titik Knot Optimal *Spline Truncated* dan Kernel 2 Titik Knot

Tabel 4.3 Skor GCV untuk 2 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Spline			Kernel	GCV	R^2
Knot(*K*)			*Bandwidth*(α)		
$t_1=K_1$ $t_1=K_2$	$t_2=K_1$ $t_2=K_2$	$t_3=K_1$ $t_3=K_2$	α α_1 α_2		
12.85 31.21	24.07 44.0	246911 425013.5	7.88 86.21	28.8	87,5%
12.85 28.46	24.07 41.0	246911 398298.2	7.84 85.74	28.8	
12.85 29.38	24.07 42.0	246911. 407203.3	7.85 85.90	28.9	
12.85 27.54	24.07 40.0	246911. 389393.1	7.82 85.58	28.9	
12.85 31.21	24.07 44.0	246911. 425013.5	7.88 86.05	29.0	

Ringkasan perhitungan pemilihan titik knot dan nilai GCV diberikan pada Tabel 4.3. Berdasarkan Tabel 4.3 dapat diketahui nilai GCV minimum adalah 31,92142. Nilai GCV minimum ini memberikan titik knot:1.

x_1 = 12.85 dan 31.21 (untuk variabel persentase persalinan yang dilakukan dengan bantuan non medis).

x_2 = 24.07 dan 44.08 (untuk variabel persentase wanita perkawinan pertama kurang dari 17 tahun).

x_3 = 246911.4 dan 425013.5 (untuk variabel rata-rata jumlah pengeluaran rumah tangga perkapita sebulan).

Pemilihan model campuran regresi nonparametrik spline truncated dan kernel terbaik ditentukan berdasarkan titik-titik knot yang optimum yang didapatkan dari nilai GCV yang paling kecil. Berikut ini adalah estimasi model campuran regresi nonparametrik spline truncated dan kernel dengan 2 titik knot :

$$\begin{aligned}\hat{y} &= \hat{\beta}_{10} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}\left(x_{1i} - k_{11}\right)_+ + \hat{\beta}_{13}\left(x_{1i} - k_{12}\right)_+ + \hat{\beta}_{20} + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}\left(x_{2i} - k_{22}\right)_+ \\ &+ \hat{\beta}_{23}\left(x_{2i} - k_{22}\right)_+ - \hat{\beta}_{30} + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}\left(x_{3i} - k_{31}\right)_+ + \hat{\beta}_{33}\left(x_{3i} - k_{32}\right)_+ \\ &+ \sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_1}K\left(\frac{t_1 - t_{1i}}{\alpha_1}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_1}K\left(\frac{t_1 - t_{1i}}{\alpha_1}\right)}\right)y_i + \sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_2}K\left(\frac{t_2 - t_{2i}}{\alpha_2}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_2}K\left(\frac{t_2 - t_{2i}}{\alpha_2}\right)}\right)y_i\end{aligned}$$

$$\begin{aligned}\hat{y} &= 1,0292 + 8,3742x_{1i} + 1,0292\left(x_{1i} - 12,85\right)_+ + 3,9799\left(x_{1i} - 31,21\right)_+ \\ &+ 1,0292 - 8,9918x_{2i} - 4,7502\left(x_{2i} - 24,07\right)_+ - 6,6493\left(x_{2i} - 44,08\right)_+ \\ &+ 2,8409 - 9,5796x_{3i} + 1,5121\left(x_{3i} - 246911,4\right)_+ - 1,0907\left(x_{3i} - 425013,5\right)_+ \\ &+ \sum_{i=1}^{38}\left(\frac{\frac{1}{7,8859}K\left(\frac{t_1 - t_{1i}}{7,8859}\right)}{\sum_{j=1}^{38}\frac{1}{7,8859}K\left(\frac{t_1 - t_{1i}}{7,8859}\right)}\right)y_i + \sum_{i=1}^{38}\left(\frac{\frac{1}{86,2140}K\left(\frac{t_2 - t_{2i}}{86,2140}\right)}{\sum_{j=1}^{38}\frac{1}{86,2140}K\left(\frac{t_2 - t_{2i}}{86,2140}\right)}\right)y_i\end{aligned}$$

4.3.3.3.Pemilihan Titik Knot Optimal *Spline* Linier 3 Titik Knot

Setelah diperoleh satu titik knot dan dua titik knot, kemudian dilanjutkan dengan tiga titik knot yang hasil perhitungannya pada Nilai GCV minimum ini memberikan titik knot:

Tabel 4.4 Skor GCV untuk 3 *Knot* pada Model campuran spline truncated dan kernel pada provinsi Jawa Timur

Spline			Kernel	GCV	R^2
Knot(*K*)			*Bandwidth*(α)		
$t_1 = K_1$ $t_1 = K_2$ $t_1 = K_3$	$t_2 = K_1$ $t_2 = K_2$ $t_2 = K_3$	$t_3 = K_1$ $t_3 = K_2$ $t_3 = K_3$	α α_1 α_2		
11,01 30,29 31,21	22,07 43,08 44,08	229101 416108 425013	9,430956 102,1399	27,47449	88,97%

Lanjutan tabel 4.4

Spline			Kernel	GCV	R^2
Knot(*K*)			*Bandwidth*(α)		
$t_1 = K_1$ $t_1 = K_2$ $t_1 = K_3$	$t_2 = K_1$ $t_2 = K_2$ $t_2 = K_3$	$t_3 = K_1$ $t_3 = K_2$ $t_3 = K_3$	α α_1 α_2		
10,09 30,29 31,21	21,07 43,08 44,08	220196 416108 425013	8,755214 95,72876	27,65107	88,97%
11,93 30,29 31,21	23,08 43,08 44,08	238006 416108 425013	9,894156 108,1948	27,65179	
12,85 29,38 31,21	24,07 43,08 44,08	246911 416108 425013	10,41575 113,9037	28,01197	
11,01 29,38 31,21	22,07 42,08 44,08	407203 416108 425013	9,32701 101,9872	28,21371	

Ringkasan perhitungan pemilihan titik knot dan nilai GCV diberikan pada Tabel 4.3. Berdasarkan Tabel 4.3 dapat diketahui nilai GCV minimum adalah 27.47449. Nilai GCV minimum ini memberikan titik knot:

x_1 = 11.01, 30.29 dan 31.21 (untuk variabel persentase persalinan yang dilakukan dengan bantuan non medis)

x_2 = 22.07, 43.08 dan 44.08 (untuk variabel persentase wanita perkawinan pertama kurang dari 17 tahun)

x_3 = 229101.2, 416108.4 dan 425013.5 (untuk variabel rata-rata jumlah pengeluaran rumah tangga perkapita sebulan) .

Pemilihan model campuran regresi nonparametrik spline truncated dan kernel terbaik ditentukan berdasarkan titik-titik knot yang optimum yang didapatkan dari nilai GCV yang paling kecil. Berikut ini adalah estimasi model campuran regresi nonparametrik spline truncated dan kernel dengan 3 titik knot :

$$\begin{aligned}\hat{y} = {} & \hat{\beta}_{10} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}\left(x_{1i} - k_{11}\right)_+ + \hat{\beta}_{13}\left(x_{1i} - k_{12}\right)_+ + \hat{\beta}_{14}\left(x_{1i} - k_{13}\right)_+ + \hat{\beta}_{20} \\ & + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}\left(x_{2i} - k_{21}\right)_+ + \hat{\beta}_{23}\left(x_{2i} - k_{22}\right)_+ + \hat{\beta}_{14}\left(x_{1i} - k_{13}\right)_+ - \hat{\beta}_{30} \\ & + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}\left(x_{3i} - k_{31}\right)_+ + \hat{\beta}_{33}\left(x_{3i} - k_{32}\right)_+ + \hat{\beta}_{34}\left(x_{3i} - k_{33}\right)_+\end{aligned}$$

$$+\sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_1}K\left(\frac{t_1-t_{1i}}{\alpha_1}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_1}K\left(\frac{t_1-t_{1i}}{\alpha_1}\right)}\right)y_i+\sum_{i=1}^{R}\left(\frac{\frac{1}{\alpha_2}K\left(\frac{t_2-t_{2i}}{\alpha_2}\right)}{\sum_{j=1}^{R}\frac{1}{\alpha_2}K\left(\frac{t_2-t_{2i}}{\alpha_2}\right)}\right)y_i$$

$$\begin{aligned}\hat{y} &= 0,0077106+0,68103x_{1i}+0,0077106(x_{1i}-11,01)_+\\ &\quad +0,33473(x_{1i}-30,29)_+ +0,0077106(x_{1i}-31,21)_+\\ &\quad +0,000080600-0,11722x_{2i}-0,55378(x_{2i}-22,07)_+\\ &\quad -0,55484(x_{2i}-43,08)_+ +0,14338(x_{1i}-44,08)_+\\ &\quad -0,080007-0,081938x_{3i}+0,000098303(x_{3i}-229101)_+\\ &\quad +0,0011259(x_{3i}-416108,4)_+ -0,0012371(x_{3i}-425013,5)_+\\ &\quad +\sum_{i=1}^{38}\left(\frac{\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38}\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}\right)y_i+\sum_{i=1}^{38}\left(\frac{\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38}\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}\right)y_i\end{aligned}$$

Untuk analisis lebih lanjut, knot yang digunakan dalam pemodelan adalah knot dengan nilai GCV terkecil dari semua knot yang dihasilkan. Tabel perbandingan nilai GCV minimum dapat dilihat pada Tabel 4.7 sebagai berikut :

Tabel 4.5 Perbandingan Nilai GCV minimum dari titik knot

No	Titik Knot	Nilai GCV Minimum
1.	1 Titik Knot	33.678
2.	2 Titik Knot	28.83
3.	3 Titik Knot	27.47449

Berdasarkan Tabel 4.5 terlihat bahwa nilai GCV terkecil adalah 27.47449 terdapat pada model campuran *spline truncated* dan kernel dengan tiga titik *knot*. Lokasi titik-titik knot dan bandwidth pada GCV minimum tersebut merupakan lokasi titik-titik knot dan bandwidth optimum.

4.3.3.4 Interpretasi Model Terbaik Regresi Nonparametrik Campuran spline dan kernel

Model terbaik pada tiga titik knot dapat dilihat sebagai berikut:

$$\begin{aligned}\hat{y} &= 0,0077106+0,68103x_{1i}+0,0077106(x_{1i}-11,01)_+\\ &\quad +0,33473(x_{1i}-30,29)_+ +0,0077106(x_{1i}-31,21)_+\end{aligned}$$

$$
\begin{aligned}
&+0,000080600-0,11722x_{2i}-0,55378\left(x_{2i}-22,07\right)_+\\
&-0,55484\left(x_{2i}-43,08\right)_+ +0,14338\left(x_{1i}-44,08\right)_+\\
&-0,080007-0,081938x_{3i}+0,000098303\left(x_{3i}-229101\right)_+\\
&+0,0011259\left(x_{3i}-416108,4\right)_+ -0,0012371\left(x_{3i}-425013,5\right)_+\\
&+\sum_{i=1}^{38}\left(\frac{\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38}\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}\right)y_i+\sum_{i=1}^{38}\left(\frac{\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38}\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}\right)y_i
\end{aligned}
$$

Model untuk pola data yang mengikuti kurva spline truncated dapat diinterpretasikan. Sementara itu, untuk pola data yang mengikuti kernel tidak dapat diinterpretasikan. Berikut ini adalah model masing-masing kelompok data dan interpretasinya :

1) Model untuk variabel persentase persalinan dengan bantuan non medis

Dengan mengasumsikan data selain persentase persalinan dengan bantuan non medis konstan, maka diperoleh model umumnya :

$$
\begin{aligned}
\hat{y}_i &= 0,0077106+0,68103x_{1i}+0,0077106\left(x_{1i}-11,01\right)_+\\
&\quad +0,33473\left(x_{1i}-30,29\right)_+ +0,0077106\left(x_{1i}-31,21\right)_+
\end{aligned}
$$

dimana:

$$
\begin{aligned}
c_1 &= 0,000080600-0,11722x_{2i}-0,55378\left(x_{2i}-22,07\right)_+\\
&\quad -0,55484\left(x_{2i}-43,08\right)_+ +0,14338\left(x_{1i}-44,08\right)_+\\
&\quad -0,080007-0,081938x_{3i}+0,000098303\left(x_{3i}-229101\right)_+\\
&\quad +0,0011259\left(x_{3i}-416108,4\right)_+ -0,0012371\left(x_{3i}-425013,5\right)_+\\
&\quad +\sum_{i=1}^{38}\left(\frac{\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38}\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}\right)y_i+\sum_{i=1}^{38}\left(\frac{\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38}\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}\right)y_i
\end{aligned}
$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

$$
\hat{y}=\begin{cases}
0,0077106+0,68103x_{1i}+c_1, & x_{1i}\le 11,01\\
-0,077183+0,6887406x_{1i}+c_1, & 11,01<x_{1i}\le 30,29\\
-10,2161547+0,7222136x_{1i}+c_1, & 30,29<x_{1i}\le 31,21\\
-11,4568025+0,7299242x_{1i}+c_1, & x_{1i}>31,21
\end{cases}
$$

a. Untuk persentase persalinan yang dilakukan dengan antuan non medis kurang dari 11,01% ketika dilakukan penambahan sebesar 1% maka persentase persalinan yang dilakukan dengan antuan non medis naik sebesar 0,68103%.

b. Untuk persentase persalinan yang dilakukan dengan antuan non medis berada di rentang nilai 11,01 sampai 30,29%, ketika dilakukan penambahan sebesar 1% maka persentase persalinan yang dilakukan dengan antuan non medis naik sebesar 0,6887406%.

c. Untuk persentase persalinan yang dilakukan dengan antuan non medis berada di rentang nilai 30,29% sampai 31,21%, ketika dilakukan penambahan sebesar 1% maka persentase persalinan yang dilakukan dengan antuan non medis naik sebesar 0,7222136%.

d. Untuk persentase persalinan yang dilakukan dengan antuan non medis lebih dari 31,21%, ketika dilakukan penambahan sebesar 1% maka persentase persalinan yang dilakukan dengan antuan non medis naik sebesar 0,7299242%.

2) Model untuk variabel persentase wanita perkawinan pertama kurang dari 17 tahun

Dengan mengasumsikan data selain penggunaan benih konstan, maka diperoleh model umumnya :

$$\begin{aligned}\hat{y}_i &= 0,000080600 - 0,11722x_{2i} - 0,55378\left(x_{2i} - 22,07\right)_+ \\ &\quad - 0,55484\left(x_{2i} - 43,08\right)_+ + 0,14338\left(x_{1i} - 44,08\right)_+\end{aligned}$$

Dimana

$$\begin{aligned}c_2 &= 0,0077106 + 0,68103x_{1i} + 0,0077106\left(x_{1i} - 11,01\right)_+ \\ &\quad + 0,33473\left(x_{1i} - 30,29\right)_+ + 0,0077106\left(x_{1i} - 31,21\right)_+ \\ &\quad - 0,080007 - 0,081938x_{3i} + 0,000098303\left(x_{3i} - 229101\right)_+ \\ &\quad + 0,0011259\left(x_{3i} - 416108,4\right)_+ - 0,0012371\left(x_{3i} - 425013,5\right)_+ \\ &\quad + \sum_{i=1}^{38}\left(\frac{\frac{1}{9,3409}K\left(\frac{t_1 - t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38}\frac{1}{9,3409}K\left(\frac{t_1 - t_{1i}}{9,3409}\right)}\right)y_i + \sum_{i=1}^{38}\left(\frac{\frac{1}{102,1398}K\left(\frac{t_2 - t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38}\frac{1}{102,1398}K\left(\frac{t_2 - t_{2i}}{102,1398}\right)}\right)y_i\end{aligned}$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

$$\hat{y} = \begin{cases} 0,0000806 - 0,11722x_{2i} + c_2, & x_{2i} \le 22,07 \\ 12,221924 - 0,671x_{2i} + c_2, & 22,07 < x_{2i} \le 43,08 \\ 36,1244312 - 1,22584_{2i} + c_2, & 43,08 < x_{2i} \le 44,08 \\ 42,444621 - 1,08246x_{2i} + c_2, & x_{2i} > 44,08 \end{cases}$$

a. Untuk persentase wanita perkawinan pertama kurang dari 17 tahun kurang dari 22,07% ketika dilakukan penambahan sebesar 1% maka persalinan yang dilakukan dengan antuan non medis turun sebesar 0,11722%.
b. Untuk persentase wanita perkawinan pertama kurang dari 17 tahun berada di rentang nilai 22,07% sampai 43,08%, ketika dilakukan penambahan sebesar 1% maka persentase wanita perkawinan pertama kurang dari 17 tahun turun sebesar 0,671%.
c. Untuk persentase wanita perkawinan pertama kurang dari 17 tahun berada di rentang nilai 43,08% sampai 44,08%, ketika dilakukan penambahan sebesar 1% maka persentase wanita perkawinan pertama kurang dari 17 tahun turun sebesar 1,22584%.
d. Untuk persentase wanita perkawinan pertama kurang dari 17 tahun lebih dari dari 44,08%, ketika dilakukan penambahan sebesar 1% maka persentase wanita perkawinan pertama kurang dari 17 tahun turun sebesar 1,08%.

$$\begin{aligned} \hat{y}_i &= -0,080007 - 0,081938x_{3i} + 0,000098303(x_{3i} - 229101)_+ \\ &\quad + 0,0011259(x_{3i} - 416108,4)_+ - 0,0012371(x_{3i} - 425013,5)_+ \end{aligned}$$

Dimana

$$\begin{aligned} c_3 &= 0,0077106 + 0,68103x_{1i} + 0,0077106(x_{1i} - 11,01)_+ \\ &\quad + 0,33473(x_{1i} - 30,29)_+ + 0,0077106(x_{1i} - 31,21)_+ \\ &\quad + 0,000080600 - 0,11722x_{2i} - 0,55378(x_{2i} - 22,07)_+ \\ &\quad - 0,55484(x_{2i} - 43,08)_+ + 0,14338(x_{1i} - 44,08)_+ \\ &\quad + \sum_{i=1}^{38} \left(\frac{\frac{1}{9,3409} K\left(\frac{t_1 - t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38} \frac{1}{9,3409} K\left(\frac{t_1 - t_{1i}}{9,3409}\right)} \right) y_i + \sum_{i=1}^{38} \left(\frac{\frac{1}{102,1398} K\left(\frac{t_2 - t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38} \frac{1}{102,1398} K\left(\frac{t_2 - t_{2i}}{102,1398}\right)} \right) y_i \end{aligned}$$

Untuk interpretasi model ini digunakan fungsi seperti berikut :

$$\hat{y} = \begin{cases} -0,080007 - 0,081938x_{3i} + c_3, & x_{3i} \leq 229101 \\ -22,6006353 - 0,0818397x_{3i} + c_3, & 229101 < x_{3i} \leq 416108,4 \\ -491,097083 - 0,807138_{3i} + c_3, & 416108,4 < x_{3i} \leq 425013,5 \\ 34,7242308 - 0,0819509x_{3i} + c_3, & x_{3i} > 425013,5 \end{cases}$$

a. Untuk rata-rata jumlah pengeluaran rumah tangga perkapital sebulan kurang dari 229.101 ketika dilakukan penambahan sebesar 1% maka rata-rata jumlah pengeluaran rumah tangga perkapital seulan turun sebesar 0,081938 %.
b. Untuk rata-rata jumlah pengeluaran rumah tangga perkapital sebulan berada di rentang nilai 229.101 sampai 416.108 ketika dilakukan penambahan sebesar 1% maka rata-rata jumlah pengeluaran rumah tangga perkapital sebulan turun sebesar 0,081839%.
c. Untuk rata-rata jumlah pengeluaran rumah tangga perkapital sebulan berada di rentang nilai 416.108 sampai 425.013 ketika dilakukan penambahan sebesar 1 maka rata-rata jumlah pengeluaran rumah tangga perkapital sebulan turun sebesar 0,807%.
d. Untuk rata-rata jumlah pengeluaran rumah tangga perkapital sebulan lebih dari 425.013 ketika dilakukan penambahan sebesar 1% maka rata-rata jumlah pengeluaran rumah tangga perkapital sebulan turun sebesar 0,08195.

Dari hasil diketahui bahwa persentase persalinan yang dilakukan dengan bantuan non medis berhubungan positif terhadap angka kematian bayi. Sementara itu untuk persentase wanita perkawinan pertama kurang dari 17 tahun dan rata-rata jumlah pengeluaran rumah tangga perkapital sebulan kerja perlu digunakan secara optimal sehingga dapat menurunkan angka kematian bayi yang optimal pula.

4.4 Interval Konfidensi Provinsi Jawa Timur

Dari model *spline Truncated* dengan tiga titik *knot* , akan dibangun interval konfidensi 95% dengan persamaan sebagai berikut:

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + b^*[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj} \leq \beta_j \leq$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I - D(\alpha))\tilde{y}]_j + a^*[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1 - \alpha$$

dengan $a^* = t_{(n-m-r-1,\alpha/2)}$ dan $b^* = -t_{(n-m-r-1,\alpha/2)}$

Berdasarkan pada Tabel 4.5 dari model *spline truncated* yang diperoleh, diketahui bahwa pada tingkat signifikansi 95%, AKB di Provinsi Jawa Timur masih berada di dalam kisaran *mean*-nya. Dari tabel di bawah ini dapat dilihat batas atas dan bawahnya, yang dapat membantu pembaca khususnya pemerintah daerah dalam memprediksi angka kematian bayi yang baik adalah yang tidak lebih dan tidak kurang dari batas atas dan batas bawah konfidensi interval yang didapat.

Table 4.6 konfidensi interval Provinsi Jawa Timur

		Batas atas	$\hat{\beta}$	Batas bawah
x_1	β_{10}	0.0122	0.0077106	0.0031326
	β_{11}	0.9954	0.68103	0.36659
	β_{12}	0.0122	0.0077106	0.0031326
	β_{13}	0.5459	0.33473	0.12349
	β_{14}	0.0122	0.0077106	0.0031326
x_2	β_{20}	-0.00005	-0.00008060	-0.00010802
	β_{21}	0.08141	-0.1172	-0.31585
	β_{22}	-0.1625	-0.5537	-0.94502
	β_{23}	-0.16267	-0.5484	-0.93419
	β_{24}	0.23285	0.14338	0.053923
x_3	β_{30}	0.19338	-0.08000	-0.35340
	β_{31}	0.17393	-0.08193	-0.33781
	β_{32}	0.00015	0.000098303	0.000037
	β_{33}	0.00209	0.0011259	0.000161
	β_{34}	-0.00023	-0.001237	-0.002235

Berdasarkan tabel 4.6 dapat di buat kesimpulan seagai berikut:

a. Untuk variabel x_1 (persentase persalinan yang dilakukan dengan bantuan non medis) memuat parameter $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$ dan β_{14}. Berdasarkan tabel 4.6 dapat dilihat bahwa $\beta_{10} \neq 0, \beta_{11} \neq 0, \beta_{12} \neq 0, \beta_{13} \neq 0$ dan $\beta_{14} \neq 0$ artinya parameter signifikan terhadap angka kematian bayi (AKB).

b. Untuk variabel x_3 (persentase wanita perkawinan pertama kurang dari 17 tahun) memuat parameter $\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}$ dan β_{24}. Berdasarkan tabel 4.6 dapat dilihat bahwa $\beta_{20} \neq 0,\ \beta_{21} = 0,\ \beta_{22} \neq 0,\ \beta_{23} \neq 0$ dan $\beta_{24} \neq 0$ artinya parameter signifikan terhadap angka kematian bayi (AKB) karena paling tidak ada satu parameter yang tidak memuat nol walaupun ada satu parameter yang memuat nol.

c. Untuk variabel x_2 (rata-rata jumlah pengeluaran rumah tangga perkapital sebulan) memuat parameter $\beta_{30}, \beta_{31}, \beta_{32}, \beta_{33}$ dan β_{34}. Berdasarkan tabel 4.6 dapat dilihat bahwa $\beta_{30} = 0,\ \beta_{31} = 0,\ \beta_{32} \neq 0,\ \beta_{33} \neq 0$ dan $\beta_{14} = 0$ artinya parameter signifikan terhadap angka kematian bayi (AKB) walaupun ada dua parameter yang memuat nol karena paling tidak ada satu parameter yang tidak memuat nol.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan pada bab 4 dapat diambil kesimpulan sebagai berikut:

1. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi diketahui.

   Interval Konfidensi untuk parameter β_j diberikan oleh:

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j - Z_{\alpha/2}[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j + Z_{\alpha/2}[\sqrt{\sigma^2\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1-\alpha$$

2. Interval konfidensi terpendek untuk parameter model komponen spline truncated variansi tidak ketahui.

   Interval Konfidensi untuk parameter β diberikan oleh:

$$P[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j - t_{(n-m-r-1,\alpha/2)}[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj} \le \beta_j \le$$
$$[[X(\tilde{\lambda})^T X(\tilde{\lambda})]^{-1} X(\tilde{\lambda})^T (I-D(\alpha))\tilde{y}]_j + t_{(n-m-r-1,\alpha/2)}[\sqrt{MSE\omega_j(\tilde{\lambda},\alpha)}]_{jj}] = 1-\alpha$$

   dengan $a^* = t_{(n-m-r-1,\alpha/2)}$ dan $b^* = -t_{(n-m-r-1,\alpha/2)}$

3. Interval Konfidensi Provinsi Jawa Timur

   Berdasarkan nilai GCV terkecil diperoleh model *spline Truncated* dengan tiga titik *knot* yang memiliki nilai GCV yaitu 27.47449 dan dimana nilai MSE dan R^2 berturut-turut adalah 4,11 dan 88,97%. Modelnya dapat dituliskan sebagai berikut.

$$\begin{aligned}
\hat{y} = {} & 0,0077106 + 0,68103x_{1i} + 0,0077106(x_{1i} - 11,01)_+ \\
& + 0,33473(x_{1i} - 30,29)_+ + 0,0077106(x_{1i} - 31,21)_+ \\
& + 0,000080600 - 0,11722x_{2i} - 0,55378(x_{2i} - 22,07)_+ \\
& - 0,55484(x_{2i} - 43,08)_+ + 0,14338(x_{1i} - 44,08)_+ \\
& - 0,080007 - 0,081938x_{3i} + 0,000098303(x_{3i} - 229101)_+ \\
& + 0,0011259(x_{3i} - 416108,4)_+ - 0,0012371(x_{3i} - 425013,5)_+
\end{aligned}$$

$$+\sum_{i=1}^{38}\left(\frac{\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}{\sum_{j=1}^{38}\frac{1}{9,3409}K\left(\frac{t_1-t_{1i}}{9,3409}\right)}\right)y_i+\sum_{i=1}^{38}\left(\frac{\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}{\sum_{j=1}^{38}\frac{1}{102,1398}K\left(\frac{t_2-t_{2i}}{102,1398}\right)}\right)y_i$$

Interval konfidensi 95% variabel model dapat ditulis sebagai berikut:

a. Untuk variabel x_1 (persentase persalinan yang dilakukan dengan bantuan non medis) memuat parameter $\beta_{10},\beta_{11},\beta_{12},\beta_{13}$ dan β_{14} dimana semua parameternya tidak memuat nol artinya parameter signifikan terhadap angka kematian bayi (AKB).

b. Untuk variabel x_3 (persentase wanita perkawinan pertama kurang dari 17 tahun) memuat parameter $\beta_{20},\beta_{21},\beta_{22},\beta_{23}$ dan β_{24} dimana parameter $\beta_{20}\neq 0,\ \beta_{21}=0,\ \beta_{22}\neq 0,\ \beta_{23}\neq 0$ dan $\beta_{24}\neq 0$ yang artinya parameter signifikan terhadap angka kematian bayi (AKB) karena paling tidak ada satu parameter yang tidak memuat nol.

c. Untuk variabel x_2 (rata-rata jumlah pengeluaran rumah tangga perkapital sebulan) memuat parameter $\beta_{30},\beta_{31},\beta_{32},\beta_{33}$ dan β_{34}. Dimana parameter $\beta_{30}=0,\ \beta_{31}=0,\ \beta_{32}\neq 0,\ \beta_{33}\neq 0$ dan $\beta_{14}=0$ yang artinya parameter signifikan terhadap angka kematian bayi (AKB) walaupun ada dua parameter yang memuat nol karena paling tidak ada satu parameter yang tidak memuat nol.

5.2 Saran

Berdasarkan hasil penelitian yang diperoleh, saran yang dapat diberikan penulis adalah sebagai berikut:

1. Bagi pemerintah karena variabel-variabel yang mempengaruhi adalah x_1 (persentase persalinan dengan bantuan non medis), x_2 (persentase wanita perkawinan pertama kurang dari 17 tahun) dan x_3 (rata-tata jumlah pengeluaran rumah tangga perkapital sebulan) maka untuk menurunkan angka kematian bayi perlu mempengaruhi variabel-variabel tersebut.
2. Untuk penelitian lebih lanjut dapat pula menggunakan pengujian hipotesis untuk variabel-variabel berpengaruh terhadap angka kematian bayi.

DAFTAR PUSTAKA

Aulele, S.N., (2010), *Model Geographically Weighted Regression (Studi Kasus : Jumlah Kematian Bayi di Propinsi Jawa Tengah Tahun 2007)*, Thesis, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.

Aydin, Dursun. 2007. *A Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel Regression*. World Academy of Science, Engineering and Technology, 36, 253-257, Turkey.

Badan Pusat Statistika, (2011), *Hasil Survie Sosial Ekonomi Nasional Tahun 2011 Propinsi Jawa Timur*, BPS, Surabaya

Budiantara, I.N., (2000), *Metode U, GML, CV dan GCV Dalam Regresi Nonparametrik Spline*, Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI), 6, 285-290.

Budiantara, I.N., (2006), *Model Spline dengan Knots Optimal*. Jurnal Natural FMIPA Universitas Jember, Jember.

Eubank, R., (1988), *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York.

Eubank, R.L., 1999, *Spline Smoothing and Nonparametric Regression 2nd Edition*, Marcel Deker, New York.

Hardle, W., (1990), *Applied Nonparametric Regression*, Cambrige University Press, Boston.

KemenKes, 2012, *Data Informasi Kesehatan Provinsi Jawa Timur*, Kementrian Kesehatan Republik Indonesia, Jakarta.

Laome, L., (2008), *Estimasi Parameter Pada Regresi Semiparametrik Untuk Data Longitudinal*. http://jurnal.unhalu.ac.id. Tanggal akses : 5 Oktober 2011.

Puspitasari, Elsha., (2012), *Model Regresi Spline Knot Optimal Untuk Mengetahui Faktor-Faktor Yang Mempengaruhi Jumlah Kematian Bayi di Jawa Timur*, Skripsi, Jurusan Matematika, FMIPA UNESA, Surabaya.

Sukarsa, K.G., (2012), *Estimator Kernel Dalam Model Regresi Nonparametrik*. Jurnal Matematika Vol. 2 No. 1, Juni 2012. ISSN : 1693-1394

Tripena, (2011), *Penentuan Model Regresi Spline Terbaik*. Prosiding Seminar Nasional Statistika Universitas Diponegoro 2011. ISBN: 978-979-097-142-4

Wasono, (2014), *Analisis Model Regresi Nonparametrik Multivariabel Heteroskedastik Spline*. Thesis, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.

Lampiran 1. Data Angka Kematian Bayi Tahun 2011 dan Faktor-Faktor yang Diduga Mempengaruhinya

No	Kabupaten /Kota	y	t_1	t_2	t_3	t_4	t_5
1	KAB. PACITAN	22,93	5,25	18,1	Rp 185.891	90,45	80
2	KAB. PONOROGO	27,32	0	20,66	Rp 164.906	100	93
3	KAB. TRENGGALEK	21,85	4,92	22,51	Rp 191.480	95,06	91
4	KAB. TULUNGAGUNG	22,27	2,06	21,81	Rp 217.796	100	104
5	KAB. BLITAR	23,71	2,5	22	Rp 206.599	88,87	97
6	KAB. KEDIRI	29,07	3,6	17,03	Rp 202.117	100	121
7	KAB. MALANG	30,75	8,28	29,79	Rp 223.389	95,09	144
8	KAB. LUMAJANG	8,55	9,39	33,67	Rp 159.745	95,25	79
9	KAB. JEMBER	56,45	27,85	38,89	Rp 174.956	97,34	189
10	KAB. BANYUWANGI	35,04	9,88	33,36	Rp 224.662	100	156
11	KAB. BONDOWOSO	54,35	20,16	59,09	Rp 145.179	92,52	90
12	KAB. SITUBONDO	54,6	19,41	56,98	Rp 156.886	92,58	80
13	KAB. PROBOLINGGO	64,19	22,66	55,79	Rp 156.013	86,27	123
14	KAB. PASURUAN	51,62	10,85	31,62	Rp 165.302	90,85	109
15	KAB. SIDOARJO	23,88	0	10,07	Rp 355.439	90,44	95
16	KAB. MOJOKERTO	25,57	1,7	18,89	Rp 225.149	88,43	90
17	KAB. JOMBANG	27,03	0,35	20,65	Rp 271.277	93,33	114
18	KAB. NGANJUK	31,45	1,48	19,79	Rp 193.273	95,44	108
19	KAB. MADIUN	31,35	2,52	23,59	Rp 197.587	83,92	86
20	KAB. MAGETAN	23,21	0	22,81	Rp 220.940	95,21	82
21	KAB. NGAWI	27,46	1,36	25,42	Rp 164.244	91,74	88
22	KAB. BOJONEGORO	38,89	8,44	34,33	Rp 169.255	95,86	110
23	KAB. TUBAN	34,84	5,53	31,22	Rp 170.502	97,64	90
24	KAB. LAMONGAN	34,02	0,95	36,37	Rp 211.199	90,33	146
25	KAB. GRESIK	23,46	0,76	21,98	Rp 269.509	86,04	115
26	KAB. BANGKALAN	54,22	33,12	30,04	Rp 130.983	90,5	91
27	KAB. SAMPANG	55,11	44,99	42,75	Rp 124.892	100	73
28	KAB. PAMEKASAN	51,66	14,32	40,89	Rp 129.818	98,48	69
29	KAB. SUMENEP	48,47	34,82	45,55	Rp 122.240	90,95	103
30	KOTA KEDIRI	25,1	0,8	13,03	Rp 348.443	100	42
31	KOTA BLITAR	20,02	0	10,72	Rp 349.553	100	24
32	KOTA MALANG	25,26	0,64	11,47	Rp 467.678	97,18	55
33	KOTA PROBOLINGGO	25,6	7,53	28,89	Rp 307.279	88,27	28
34	KOTA PASURUAN	41,31	4,34	18,87	Rp 424.897	87,34	37

35	KOTA MOJOKERTO	22,21	0,68	11,86	Rp 415.442	100	24

Lanjutan Lampiran 1. Data Angka Kematian Bayi Tahun 2011 dan Faktor-Faktor yang Diduga Mempengaruhinya

No	Kabupaten /Kota	y	t_1	t_2	t_3	t_4	t_5
36	KOTA MADIUN	23,43	2,13	11,31	Rp 385.831	93,98	29
37	KOTA SURABAYA	23,35	1,92	11,76	Rp 558.590	95,23	156
38	KOTA BATU	29,27	0,63	22,59	Rp 305.292	95,6	13

Keterangan :

y = Angka Kematian Bayi (per 1.000 kelahiran)

t_1 = persentase persalinan yang dilakukan dengan bantuan non medis (%)

t_2 = persentase wanita perkawinan pertama kurang dari 17 tahun (%)

t_3 = rata-rata jumlah pengeluaran rumah tangga perkapita sebulan (Rupiah)

t_4 = persentase bayi usia 0-11 bulan di beri ASI (%)

t_5 = jumlah sarana kesehatan

Lampiran 2

Program Interval Konfidensi *Spline* 1 Knots

```
library(pracma)
data=read.csv('d:\\datasplineinterval.csv',sep=';',header=FALSE)
y=data[,3]
xs=as.matrix(data[,c(4:6)])
xk=as.matrix(data[,c(7,8)])
ns=ncol(xs)
nk=ncol(xk)
n=nrow(xs)
identitas=diag(1,n,n)
mat1n=matrix(1,n)
mat1nn=matrix(1,n,n)
nknot=50
nband=50
knot=matrix(0,nknot,ns)
band=matrix(0,nknot,nk)
for (i in 1:ns)
{
  a=seq(min(xs[,i]),max(xs[,i]),length.out=50)
  knot[,i]=a
}
knot=knot[2:(nknot-1),]
for (i in 1:nk)
{
  a=seq(0.01,max(xk[,i])-min(xk[,i])+0.01,length.out=50)
```

```
 band[,i]=a
}
band=band[2:(nband-1),]
nknot=nrow(knot)
nband=nknot
# desain matriks Z(lambda) pada spline
X=cbind(1,xs[,1],1,xs[,2],1,xs[,3])
MSE=matrix(0,nrow=nknot)
GCV=matrix(0,nrow=nknot)
for (i in 1:nknot)
{
 Xs=cbind(pmax(0,xs[,1]-knot[i,1]),pmax(0,xs[,2]-knot[i,2]),pmax(0,xs[,3]-knot[i,3]))
 Z=cbind(X,Xs)

 Dkernel=0    #mendefinisikan agar untuk setiap perulangan, nilainya 0
 for (j in 1:nk)
 {
  xkdiag=diag(xk[,j])
  V=mat1nn%*%xkdiag
  zz=(t(V)-V)/band[i,j]
  K=1/sqrt(2*pi)*exp(-1/2*zz^2)             # fungsi kernel gaussian
  K.Z=(1/band[i,j])*K
  W.penyebut=diag(c(1/n*K.Z%*%mat1n))%*%mat1nn
  Dker=1/n*K.Z/W.penyebut                                     # penimbang V(phi).1
  Dkernel=Dkernel+Dker        #nilai kernel untuk setiap variabel
 }
 D_kernel=Dkernel/nk
```

```
  B=pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)%*%y
  Sg=Z%*%pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)
  Zg=Sg+D_kernel
  yhat=Zg%*%y
  error=y-yhat
  MSE[i]=n^-1*t(error)%*%error
  db=(n^-1*sum(diag(identitas-Zg)))^2
  GCV[i]=MSE[i]/db
}

optimum=cbind(knot,band,MSE,GCV)
colnames(optimum)=c("knotx1","knotx2","knotx3","phix4","phix5","MSE","GCV")
#memberi nama variabel
GCVmin=optimum[order(optimum[,7]),] #mengurutkan nilai GCV minimum
#write.csv(GCVmin,file="e:/output1.csv")

#valid
knotoptim=GCVmin[1,1:3]
bandoptim=GCVmin[1,4:5]
Xs=cbind(pmax(0,xs[,1]-GCVmin[1,1]),pmax(0,xs[,2]-GCVmin[1,2]),pmax(0,xs[,3]-
GCVmin[1,3]))
Z=cbind(X,Xs)

Dkernel=0    #mendefinisikan agar untuk setiap perulangan, nilainya 0
for (j in 1:2)
{
  xkdiag=diag(xk[,j])
```

```
  V=mat1nn%*%xkdiag
  zz=(t(V)-V)/GCVmin[1,j+3]
  K=1/sqrt(2*pi)*exp(-1/2*zz^2)            # fungsi kernel gaussian
  K.Z=(1/GCVmin[1,j+3])*K
  W.penyebut=diag(c(1/n*K.Z%*%mat1n))%*%mat1nn
  Dker=1/n*K.Z/W.penyebut                              # penimbang V(phi).1
  Dkernel=Dkernel+Dker       #nilai kernel untuk setiap variabel
}
D_kernel=Dkernel/nk

#confiden interval
B=pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)%*%y
A=Z%*%pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)
C=A+D_kernel
yhat=C%*%y
SST=sum((y-mean(y))^2)
SSR=sum((yhat-mean(y))^2)
SSE=sum((y-yhat)^2)
MSR=SSR/(ns+nk)
MSE=SSE/(n-(ns+nk)-1)
R2=SSR/SST
e=y-yhat
expb=B
W=pinv(t(Z)%*%Z)
sigma=GCVmin[1,6] #sigma yang saya gunakan yaitu MSE dari knot terbaik
varb=sigma*W
w=diag(varb)
```

```
ztab=qnorm(0.975,0,1)
pa=expb-ztab*sqrt(w)
pb=expb+ztab*sqrt(w)
CI=cbind(pa,B,pb)
colnames(CI)=c("CI bawah", "Beta", "CI atas")
knotoptim
bandoptim
B
R2
CI
```

Lampiran 3

Program Interval Konfidensi *Spline* 2 Knots

```
library(pracma)
data=read.csv('d:\\datasplineinterval.csv',sep=';',header=FALSE)
y=data[,3]
xs=as.matrix(data[,c(4:6)])
xk=as.matrix(data[,c(7,8)])
ns=ncol(xs)
nk=ncol(xk)
n=nrow(xs)
identitas=diag(1,n,n)
mat1n=matrix(1,n)
mat1nn=matrix(1,n,n)
nknot=50
knot=matrix(0,nknot,ns)

for (i in 1:ns)
{
  a=seq(min(xs[,i]),max(xs[,i]),length.out=50)
  knot[,i]=a
}
knot=knot[2:(nknot-1),]
nknot=nrow(knot)
nkomb=(nknot*(nknot-1)/2)
knot2=matrix(0,nkomb,6)
v=1
```

```
for (i in 1:(nknot-1))
{
  for (j in (i+1):nknot)
  {
    knot2[v,]=cbind(knot[i,1],knot[j,1],knot[i,2],knot[j,2],knot[i,3],knot[j,3])
    v=v+1
  }
}
nknot=nrow(knot2)
band=matrix(0,nknot,nk)
for (i in 1:nk)
{
  a=seq(0.01,max(xk[,i])-min(xk[,i])+0.01,length.out=nknot)
  band[,i]=a
}
nband=nknot

#desain matriks Z(lambda) pada spline
X=cbind(1,xs[,1],1,xs[,2],1,xs[,3])
MSE=matrix(0,nrow=nknot)
GCV=matrix(0,nrow=nknot)
for (i in 1:nknot)
{
  Xs=cbind(pmax(0,xs[,1]-knot2[i,1]),pmax(0,xs[,1]-knot2[i,2]),pmax(0,xs[,2]-
knot2[i,3]),pmax(0,xs[,2]-knot2[i,4]),pmax(0,xs[,3]-knot2[i,5]),pmax(0,xs[,3]-knot2[i,6]))
  Z=cbind(X,Xs)
```

```
 Dkernel=0    #mendefinisikan agar untuk setiap perulangan, nilainya 0
 for (j in 1:nk)
 {
  xkdiag=diag(xk[,j])
  V=mat1nn%*%xkdiag
  zz=(t(V)-V)/band[i,j]
  K=1/sqrt(2*pi)*exp(-1/2*zz^2)              # fungsi kernel gaussian
  K.Z=(1/band[i,j])*K
  W.penyebut=diag(c(1/n*K.Z%*%mat1n))%*%mat1nn
  Dker=1/n*K.Z/W.penyebut                                    # penimbang V(phi).1
  Dkernel=Dkernel+Dker         #nilai kernel untuk setiap variabel
 }
 D_kernel=Dkernel/nk

 B=pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)%*%y
 Sg=Z%*%pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)
 Zg=Sg+D_kernel
 yhat=Zg%*%y
 error=y-yhat
 MSE[i]=n^-1*t(error)%*%error
 db=(n^-1*sum(diag(identitas-Zg)))^2
 GCV[i]=MSE[i]/db
}
optimum=cbind(knot2,band,MSE,GCV)
colnames(optimum)=c("knot1x1","knot2x1","knot1x2","knot2x2","knot1x3","knot2x3","phix4","phix5","MSE","GCV")  #memberi nama variabel
GCVmin=optimum[order(optimum[,10]),] #mengurutkan nilai GCV minimum
```

```
write.csv(GCVmin,file="e:/output2.csv")


#valid

knotoptim=GCVmin[1,1:6]

bandoptim=GCVmin[1,7:8]

Xs=cbind(pmax(0,xs[,1]-GCVmin[1,1]),pmax(0,xs[,1]-GCVmin[1,2]),pmax(0,xs[,2]-
GCVmin[1,3]),pmax(0,xs[,2]-GCVmin[1,4]),pmax(0,xs[,3]-GCVmin[1,5]),pmax(0,xs[,3]-
GCVmin[1,6]))

Z=cbind(X,Xs)


Dkernel=0    #mendefinisikan agar untuk setiap perulangan, nilainya 0

for (j in 1:2)

{

 xkdiag=diag(xk[,j])

 V=mat1nn%*%xkdiag

 zz=(t(V)-V)/GCVmin[1,j+6]

 K=1/sqrt(2*pi)*exp(-1/2*zz^2)                # fungsi kernel gaussian

 K.Z=(1/GCVmin[1,j+6])*K

 W.penyebut=diag(c(1/n*K.Z%*%mat1n))%*%mat1nn

 Dker=1/n*K.Z/W.penyebut                                        # penimbang V(phi).1

 Dkernel=Dkernel+Dker         #nilai kernel untuk setiap variabel

}

D_kernel=Dkernel/nk


#confiden interval

B=pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)%*%y

A=Z%*%pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)

C=A+D_kernel
```

```
yhat=C%*%y
SST=sum((y-mean(y))^2)
SSR=sum((yhat-mean(y))^2)
SSE=sum((y-yhat)^2)
MSR=SSR/(ns+nk)
MSE=SSE/(n-(ns+nk)-1)
R2=SSR/SST
e=y-yhat
expb=B
W=pinv(t(Z)%*%Z)
sigma=GCVmin[1,9] #sigma yang saya gunakan yaitu MSE dari knot terbaik
varb=sigma*W
w=diag(varb)
ztab=qnorm(0.975,0,1)
pa=expb-ztab*sqrt(w)
pb=expb+ztab*sqrt(w)
CI=cbind(pa,B,pb)
colnames(CI)=c("CI bawah", "Beta", "CI atas")
knotoptim
bandoptim
B
R2
CI
```

Lampiran 2

Program Interval Konfidensi *Spline* 1 Knots

```
library(pracma)

data=read.csv('d:\\datasplineinterval.csv',sep=';',header=FALSE)

y=data[,3]

xs=as.matrix(data[,c(4:6)])

xk=as.matrix(data[,c(7,8)])

ns=ncol(xs)

nk=ncol(xk)

n=nrow(xs)

identitas=diag(1,n,n)

mat1n=matrix(1,n)

mat1nn=matrix(1,n,n)

nknot=50

knot=matrix(0,nknot,ns)


for (i in 1:ns)

{

  a=seq(min(xs[,i]),max(xs[,i]),length.out=50)

  knot[,i]=a

}

knot=knot[2:(nknot-1),]

nknot=nrow(knot)

nkomb=(nknot*(nknot-1)*(nknot-2)/6)

knot3=matrix(0,nkomb,9)

v=1
```

```
for (i in 1:(nknot-2))
{
 for (j in (i+1):(nknot-1))
 {
  for (k in (j+1):nknot)
  {

knot3[v,]=cbind(knot[i,1],knot[j,1],knot[k,1],knot[i,2],knot[j,2],knot[k,2],knot[i,3],knot[j,3],
knot[k,3])
   v=v+1
  }
 }
}
nknot=nrow(knot3)
band=matrix(0,nknot,nk)
for (i in 1:nk)
{
 a=seq(0.01,max(xk[,i])-min(xk[,i])+0.01,length.out=nknot)
 band[,i]=a
}
nband=nknot


#desain matriks Z(lambda) pada spline
X=cbind(1,xs[,1],1,xs[,2],1,xs[,3])
MSE=matrix(0,nrow=nknot)
GCV=matrix(0,nrow=nknot)
for (i in 1:nknot)
{
```

```
  Xs=cbind(pmax(0,xs[,1]-knot3[i,1]),pmax(0,xs[,1]-knot3[i,2]),pmax(0,xs[,1]-
knot3[i,3]),pmax(0,xs[,2]-knot3[i,4]),pmax(0,xs[,2]-knot3[i,5]),pmax(0,xs[,2]-
knot3[i,6]),pmax(0,xs[,3]-knot3[i,7]),pmax(0,xs[,3]-knot3[i,8]),pmax(0,xs[,3]-knot3[i,9]))
  Z=cbind(X,Xs)

  Dkernel=0    #mendefinisikan agar untuk setiap perulangan, nilainya 0
  for (j in 1:nk)
  {
   xkdiag=diag(xk[,j])
   V=mat1nn%*%xkdiag
   zz=(t(V)-V)/band[i,j]
   K=1/sqrt(2*pi)*exp(-1/2*zz^2)              # fungsi kernel gaussian
   K.Z=(1/band[i,j])*K
   W.penyebut=diag(c(1/n*K.Z%*%mat1n))%*%mat1nn
   Dker=1/n*K.Z/W.penyebut                                          # penimbang V(phi).1
   Dkernel=Dkernel+Dker         #nilai kernel untuk setiap variabel
  }
  D_kernel=Dkernel/nk

  B=pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)%*%y
  Sg=Z%*%pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)
  Zg=Sg+D_kernel
  yhat=Zg%*%y
  error=y-yhat
  MSE[i]=n^-1*t(error)%*%error
  db=(n^-1*sum(diag(identitas-Zg)))^2
  GCV[i]=MSE[i]/db
}
```

```
optimum=cbind(knot3,band,MSE,GCV)

colnames(optimum)=c("knot1x1","knot2x1","knot3x1","knot1x2","knot2x2","knot3x2","kn
ot1x3","knot2x3","knot3x3","phix4","phix5","MSE","GCV")  #memberi nama variabel

GCVmin=optimum[order(optimum[,13]),] #mengurutkan nilai GCV minimum

write.csv(GCVmin,file="e:/output3.csv")


#valid

knotoptim=GCVmin[1,1:9]

bandoptim=GCVmin[1,10:11]

Xs=cbind(pmax(0,xs[,1]-GCVmin[1,1]),pmax(0,xs[,1]-GCVmin[1,2]),pmax(0,xs[,1]-
GCVmin[1,3]),pmax(0,xs[,2]-GCVmin[1,4]),pmax(0,xs[,2]-GCVmin[1,5]),pmax(0,xs[,2]-
GCVmin[1,6]),pmax(0,xs[,3]-GCVmin[1,7]),pmax(0,xs[,3]-GCVmin[1,8]),pmax(0,xs[,3]-
GCVmin[1,9]))

Z=cbind(X,Xs)


Dkernel=0    #mendefinisikan agar untuk setiap perulangan, nilainya 0

for (j in 1:2)

{

 xkdiag=diag(xk[,j])

 V=mat1nn%*%xkdiag

 zz=(t(V)-V)/GCVmin[1,j+6]

 K=1/sqrt(2*pi)*exp(-1/2*zz^2)              # fungsi kernel gaussian

 K.Z=(1/GCVmin[1,j+6])*K

 W.penyebut=diag(c(1/n*K.Z%*%mat1n))%*%mat1nn

 Dker=1/n*K.Z/W.penyebut                                         # penimbang V(phi).1

 Dkernel=Dkernel+Dker        #nilai kernel untuk setiap variabel

}

D_kernel=Dkernel/nk
```

```
#confiden interval

B=pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)%*%y

A=Z%*%pinv(t(Z)%*%Z)%*%t(Z)%*%(identitas-D_kernel)

C=A+D_kernel

yhat=C%*%y

SST=sum((y-mean(y))^2)

SSR=sum((yhat-mean(y))^2)

SSE=sum((y-yhat)^2)

MSR=SSR/(ns+nk)

MSE=SSE/(n-(ns+nk)-1)

R2=SSR/SST

e=y-yhat

expb=B

W=pinv(t(Z)%*%Z)

sigma=GCVmin[1,12] #sigma yang saya gunakan yaitu MSE dari knot terbaik

varb=sigma*W

w=diag(varb)

ztab=qnorm(0.975,0,1)

pa=expb-ztab*sqrt(w)

pb=expb+ztab*sqrt(w)

CI=cbind(pa,B,pb)

colnames(CI)=c("CI bawah", "Beta", "CI atas")

knotoptim

bandoptim

B

R2

CI
```

BIOGRAFI PENULIS

Syisliawati, Lahir di Makassar pada tanggal 03 Maret 1992. Penulis merupakan anak kedua dari tiga bersaudara dari pasangan H. Abdul Samad dan Hj.Hartati. Penulis mulai memasuki jenjang pendidikan formal pada Sekolah Dasar Negeri 18 Ujung pada tahun 1998-2003, dan pindah ke sekolah Sekolah Dasar Negeri 3 Tala pada tahun 2003-2004. Kemudian penulis melanjutkan pendidikan di Sekolah Lanjutan Tingkat Pertama Negeri 1 Ma’rang pada tahun 2004-2007. Pada tahun yang sama penulis melanjutkan pendidikan ke Sekolah Menengah Atas Negeri 1 Ma’rang pada tahun 2007-2010. Penulis kemudian melanjutkan pendidikan Sarjana (S1) di Universitas Negeri Makassar (UNM) pada Program Studi Pendidikan Matematika dan selesai pada tahun 2014. Kemudian pada tahun yang sama Penulis melanjutkan studi Pasca Sarjana (S2) di Institut Teknologi Sepuluh Nopember Program studi Statistika. Segala kritik dan Saran yang dapat dikirim melalui email : syisliailasamad123@gmail.com