TESIS - SS14 2501

METODE *UNBIASED RISK* (UBR) DAN *CROSS-VALIDATION* (CV) UNTUK PEMILIHAN TITIK KNOT OPTIMAL DALAM REGRESI NONPARAMETRIK MULTIVARIABEL SPLINE *TRUNCATED*

(Studi Kasus: Data Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah Tahun 2015)

ALVITA RACHMA DEVI
NRP. 06211650010005

DOSEN PEMBIMBING
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Dr. Vita Ratnasari, M.Si.

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

THESIS - SS14 2051

UNBIASED RISK (UBR) AND CROSS-VALIDATION (CV) METHOD FOR SELECTING OPTIMAL KNOT IN MULTIVARIABLE NONPARAMETRIC REGRESSION SPLINE TRUNCATED

(Case Study: Unemployment Rate in Central Java 2015)

ALVITA RACHMA DEVI
NRP. 06211650010005

SUPERVISOR
Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Dr. Vita Ratnasari, M.Si.

PROGRAM OF MAGISTER
DEPARTEMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTATING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018

METODE *UNBIASED RISK (UBR)* DAN *CROSS-VALIDATION (CV)* UNTUK PEMILIHAN TITIK KNOT OPTIMAL DALAM REGRESI NONPARAMETRIK MULTIVARIABEL SPLINE *TRUNCATED*
(Studi Kasus: Data Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah Tahun 2015)

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

ALVITA RACHMA DEVI
NRP. 06211650010005

Tanggal Ujian : 17 Januari 2018
Periode Wisuda : Maret 2018

Disetujui Oleh:

1. Prof.Dr.Drs. I Nyoman Budiantara, M.Si. (Pembimbing I)
   NIP.19650603 198903 1 003
2. Dr. Vita Ratnasari, S.Si., M.Si. (Pembimbing II)
   NIP.19700910 199702 2 001
3. Dr. Ismaini Zain, S.Si., M.Si. (Penguji)
   NIP. 19600525 198803 2 001
4. Dr. Suhartono, M.Sc. (Penguji)
   NIP.19710929 199512 1 001

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

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

METODE *UNBIASED RISK* (UBR) DAN *CROSS-VALIDATION* (CV) UNTUK PEMILIHAN TITIK KNOT OPTIMAL DALAM REGRESI NONPARAMETRIK MULTIVARIABEL SPLINE *TRUNCATED*

(Studi Kasus: Data Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah Tahun 2015)

Nama Mahasiswa	: Alvita Rachma Devi
NRP	: 06211650010005
Pembimbing I	: Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Pembimbing II	: Dr. Vita Ratnasari, M.Si.

ABSTRAK

Regresi nonparametrik memberikan fleksibilitas yang tinggi karena bentuk estimasi kurva regresinya menyesuaikan data. Salah satu metode regresi nonparametrik adalah spline *truncated*. Banyaknya titik knot dan lokasi dari tiap-tiap knot dalam spline *truncated* akan berpengaruh pada bentuk estimasi kurva regresi. Untuk itulah perlu dicari titik knot yang optimal untuk menghasilkan model terbaik. Ada beberapa metode yang dapat digunakan untuk memilih titik knot yang optimal, diantaranya *Unbiased Risk* (UBR) dan *Cross-Validation* (CV). Penelitian ini mengkaji metode UBR dan CV dan selanjutnya membandingkan metode UBR dan CV dalam pemilihan titik knot optimal pada data simulasi dan data Tingkat Pengangguran Terbuka (TPT) di Jawa Tengah tahun 2015. Kriteria untuk memilih model terbaik yaitu berdasarkan nilai *Mean Squared Error (MSE)* dan R^2. Simulasi dilakukan dengan membangkitkan fungsi spline *truncated* dengan eror berdistribusi normal pada kombinasi variasi ukuran sampel dan varians eror. Hasil simulasi menunjukkan metode CV menduga titik knot lebih tepat dibandingkan UBR. Hasil aplikasi pada data TPT didapatkan titik knot optimal menggunakan metode CV adalah kombinasi titik knot 2-3-2 dengan MSE sebesar 0,561 dan R^2 sebesar 86,811%. Sementara itu, dengan menggunakan metode UBR didapatkan titik knot optimal adalah tiga titik knot dengan MSE sebesar 0,996 dan R^2 sebesar 78,550%. Sehingga, baik dari hasil data simulasi maupun data riil, metode CV menduga titik knot lebih baik dibandingkan metode UBR.

Kata kunci: *Unbiased Risk* (UBR), *Cross-Validation* (CV), *Spline Truncated*, Regresi Nonparametrik, Tingkat Pengangguran Terbuka

(halaman ini sengaja dikosongkan)

UNBIASED RISK (UBR) AND CROSS-VALIDATION (CV) METHOD FOR SELECTING OPTIMAL KNOT IN MULTIVARIABLE NONPARAMETRIC REGRESSION SPLINE TRUNCATED

(Case Study: Unemployment Rate in Central Java 2015)

Name	: Alvita Rachma Devi
NRP	: 06211650010005
Supervisor	: Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Co-Supervisor	: Dr. Vita Ratnasari, M.Si.

ABSTRACT

Nonparametric regression gives better flexibility because the form of regression curve estimation adjust to data. One of nonparametric regression method is spline truncated. The number of knot and location of each knot in spline truncated affect to the form of regression curve estimation. Optimal knot is needed in order to obtain the best model. There are some methods to select optimal knot, such as Unbiased Risk (UBR) and Cross-Validation (CV). This research discuss about UBR and CV method, then compare UBR and CV for selecting optimal knot on simulation data and unemployment rate data of Central Java Province, Indonesia in 2015. Criteria for selecting the best model is based on Mean Squared Error (MSE) and R^2. The simulation is performed on a spline truncated function with error generated from normal distribution for varied sample size and error variance. The results of simulation study show that CV estimates knots better than UBR. From the result of application on unemployment rate data, the optimal knot by using CV method is on combination of 2-3-2 knot with MSE of 0,561 and R^2 of 86,111%. Meanwhile, by using UBR method the optimal knot is on three knot with MSE of 0,996 and R^2 of 78,550%. In conclusion, from results of simulation data and unemployment rate data, CV method generates a better model than UBR method.

Keywords: Unbiased Risk (UBR), Cross-Validation (CV), Spline Truncated, Nonparametric Regression, Unemployment Rate

(halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT, karena atas rahmat-Nya, tesis dengan judul "Metode *Unbiased Risk* (UBR) dan *Cross-Validation* (CV) untuk Pemilihan Titik Knot Optimal dalam Regresi Nonparametrik Multivariabel Spline *Truncated* (Studi Kasus: Data Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah Tahun 2015)" ini bisa terselesaikan. Tesis ini merupakan salah satu syarat untuk menyelesaikan pendidikan di Program Magister S2 Statistika ITS. Ada banyak pihak yang telah membantu dalam penulisan tesis ini, sehingga penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. Drs. I Nyoman Budiantara, M.Si dan ibu Dr. Vita Ratnasari, S.Si, M.Si. selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, saran dan ilmu yang bermanfaat dalam menyelesaikan tesis ini.
2. Ibu Dr. Dra. Ismaini Zain, M.Si. selaku Wakil Dekan FMKSD ITS dan dosen penguji, juga Bapak Dr. Suhartono, M.Sc. selaku Ketua Departemen Statistika ITS dan dosen penguji yang telah memberikan saran dan masukan yang menjadikan tesis ini lebih baik.
3. Bapak Dr.rer.pol. Heri Kuswanto, M.Si selaku Kaporodi Pascasarjana Statistika ITS.
4. Bapak/ibu dosen pengajar di Departemen Statistika ITS, terima kasih atas semua ilmu berharga yang diberikan.
5. Kedua orangtua yang penulis sayangi dan hormati, yang tanpa doa dan dukungannya penulis tidak akan bisa mencapai titik ini.
6. Risky Frasetio W.P. dalam membantu pengerjaan program R untuk tesis ini.
7. Teman-teman seperjuangan S2 Statistika ITS, terima kasih atas kebersamaan dan motivasinya selama mengejar ilmu di Departemen Statistika ITS.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, kritik maupun saran sangat penulis harapkan demi perbaikan tesis ini. Penulis berharap tesis ini bermanfaat dan memperluas wawasan keilmuan pembacanya.

Surabaya, Januari 2018

Penulis

(halaman ini sengaja dikosongkan)

DAFTAR ISI

LEMBAR PENGESAHAN v
ABSTRAK vii
ABSTRACT ix
KATA PENGANTAR xi
DAFTAR ISI xiii
DAFTAR TABEL xv
DAFTAR GAMBAR xvii
DAFTAR NOTASI xix
BAB 1 PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 5
1.3 Tujuan 6
1.4 Manfaat 6
1.5 Batasan Masalah 6
BAB 2 TINJAUAN PUSTAKA 7
2.1 Analisis Regresi 7
2.2 Regresi Spline *Truncated* 8
2.3 Regresi Spline *Truncated* Multivariabel 10
2.4 Pemilihan Titik Knot Optimal 10
2.5 Pengujian Parameter Model 11
2.6 *Mean Squared Error (MSE)* dan Koefisien Determinasi 12
2.7 Nilai Ekspektasi dari Bentuk Kuadratik 13
2.8 Tingkat Pengangguran Terbuka (TPT) 14
2.9 Faktor-faktor yang Mempengaruhi TPT 15
BAB 3 METODOLOGI PENELITIAN 17
3.1 Sumber Data dan Variabel Penelitian 17
3.2 Struktur Data Penelitian 18
3.3 Langkah-Langkah Penelitian 18
3.3.1 Kajian Teoritis 18
3.3.2 Studi Simulasi 20
3.3.3 Aplikasi Pada Data Tingkat Pengangguran Terbuka 21

BAB 4 HASIL DAN PEMBAHASAN 23
4.1 Kajian Teori *Unbiased Risk* dan *Cross-Validation* pada Regresi Nonparametrik Spline *Truncated* Multivariabel 23
4.1.1 Regresi Nonparametrik Spline *Truncated* Multivariabel 23
4.1.2 Kajian Metode *Unbiased Risk* (UBR) 27
4.1.3 Kajian Metode *Cross-Validation* (CV) 29
4.2 Studi Simulasi 30
4.3 Aplikasi *Unbiased Risk* (UBR) dan *Cross-Validation* (CV) pada Data Tingkat Pengangguran Terbuka di Jawa Tengah 32
4.3.1 Analisis Deskriptif 32
4.3.2 Pemodelan Tingkat Pengangguran Terbuka di Jawa Tengah Menggunakan Regresi Nonparametrik Spline *Truncated* Linier 37
4.3.2.1 Pemilihan Titik Knot Optimal dengan Satu Titik Knot 37
4.3.2.2 Pemilihan Titik Knot Optimal dengan Dua Titik Knot 38
4.3.2.3 Pemilihan Titik Knot Optimal dengan Tiga Titik Knot 40
4.3.2.4 Pemilihan Titik Knot Optimal dengan Kombinasi Satu, Dua dan Tiga Titik Knot 41
4.3.3 Pemilihan Model Terbaik 43
4.3.4 Pengujian Signifikansi Parameter Secara Simultan 47
4.3.5 Pengujian Signifikansi Parameter Secara Parsial 48
4.3.6 Pemilihan Model Terbaik Menggunakan 4 Variabel Prediktor 50
4.3.7 Pemilihan Model Terbaik Menggunakan 3 Variabel Prediktor 53
4.3.8 Interpretasi Model Regresi Nonparametrik Spline *Truncated* 56
BAB 5 KESIMPULAN DAN SARAN 59
5.1 Kesimpulan 59
5.2 Saran 60
DAFTAR PUSTAKA 61
LAMPIRAN 65
BIOGRAFI PENULIS 85

DAFTAR TABEL

Tabel 2.1 Analisis Ragam (ANOVA) 12
Tabel 3.1 Variabel Penelitian 17
Tabel 3.2 Definisi Operasional Variabel Penelitian 17
Tabel 3.3 Struktur Data Penelitian 18
Tabel 4.1 Rata-rata Estimasi Titik Knot dengan CV dan UBR 31
Tabel 4.2 Statistik Deskriptif Variabel Respon dan Variabel Prediktor 33
Tabel 4.3 Tabel Hasil Uji Linearitas dengan Uji Terasvirta 36
Tabel 4.4 Nilai UBR untuk Spline Linier Satu Knot 37
Tabel 4.5 Nilai CV untuk Spline Linier Satu Knot 38
Tabel 4.6 Nilai UBR untuk Spline Linier Dua Knot 39
Tabel 4.7 Nilai CV untuk Spline Linier Dua Knot 39
Tabel 4.8 Nilai UBR untuk Spline Linier Tiga Knot 40
Tabel 4.9 Nilai CV untuk Spline Linier Tiga Knot 41
Tabel 4.10 Nilai UBR untuk Spline Linier Kombinasi Knot 42
Tabel 4.11 Nilai CV untuk Spline Linier Kombinasi Knot 42
Tabel 4.12 Perbandingan Nilai UBR dan CV Minimum 43
Tabel 4.13 Perbandingan Nilai MSE dan R^2 Metode UBR dan CV 43
Tabel 4.14 Perbandingan Nilai Estimasi TPT ($\hat{Y}$) Metode UBR dan CV 44
Tabel 4.15 ANOVA Uji Simultan Metode CV 48
Tabel 4.16 ANOVA Uji Simultan Metode UBR 48
Tabel 4.17 Hasil Uji Signifikansi Parameter Parsial Metode CV 49
Tabel 4.18 Hasil Uji Signifikansi Parameter Parsial Metode UBR 49
Tabel 4.19 Nilai UBR Spline Linier Knot 1, 2, 3 dan Kombinasi 51
Tabel 4.20 Nilai CV Spline Linier Knot 1, 2, 3 dan Kombinasi 51
Tabel 4.21 Perbandingan Nilai MSE dan R^2 Metode UBR dan CV 51
Tabel 4.22 ANOVA Uji Simultan Metode CV 52
Tabel 4.23 Hasil Uji Signifikansi Parameter Parsial Metode CV 52
Tabel 4.24 Nilai UBR Spline Linier Knot 1, 2, 3 dan Kombinasi 53
Tabel 4.25 Nilai CV Spline Linier Knot 1, 2, 3 dan Kombinasi 54

Tabel 4.26 Perbandingan Nilai MSE dan R^2 Metode UBR dan CV 54
Tabel 4.27 ANOVA Uji Simultan Metode CV 55
Tabel 4.28 Hasil Uji Signifikansi Parameter Parsial Metode CV 55

DAFTAR GAMBAR

Gambar 4.1 Kabupaten/Kota di Jawa Tengah Berdasarkan Tingkat Pengangguran Terbuka 33
Gambar 4.2 *Scatter Plot* antara TPT Jawa Tengah Tahun 2015 dengan Variabel-Variabel yang Diduga Mempengaruhi..... 35
Gambar 4.3 Plot Nilai Estimasi TPT ($\hat{Y}$) Dan Nilai Aktual TPT (Y) Menggunakan UBR 45
Gambar 4.4 Plot Nilai Estimasi TPT ($\hat{Y}$) Dan Nilai Aktual TPT (Y) Menggunakan CV 46
Gambar 4.5 Plot Nilai Estimasi TPT ($\hat{Y}$) Dan Nilai Aktual TPT (Y) Menggunakan CV dan UBR 47

(halaman ini sengaja dikosongkan)

DAFTAR NOTASI

a	:	nilai minimum dari x
$\mathbf{A}$	:	matriks simetris yang berisi konstanta
$\mathbf{A(K)}$	:	matriks perkalian $\mathbf{X}(\mathbf{K})(\mathbf{X}'(\mathbf{K})\,\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}'(\mathbf{K})$ yang bersifat semi-definit positif
$\mathbf{A}'\mathbf{(K)}$	:	*transpose* dari $\mathbf{A(K)}$
$\propto$	:	taraf signifikansi
b	:	nilai maksimum dari x
β_0	:	konstanta parameter
β_j	:	parameter/koefisien untuk variabel ke- j
$\hat{\beta}_{jl}$	:	estimasi parameter/koefisien pada variabel ke- j dan orde ke- l
$\underset{\sim}{\beta}$	:	vektor koefisien regresi spline
$\underset{\sim}{\hat{\beta}}'$	:	*transpose* dari estimasi $\underset{\sim}{\beta}$
CV(K)	:	fungsi *cross-validation* dengan titik knot K
$\underset{\sim}{\delta}_1$	:	vektor koefisien regresi untuk fungsi polinomial pada regresi spline
$\underset{\sim}{\delta}_2$	:	vektor koefisien regresi untuk fungsi *truncated* pada regresi spline
ε_i	:	residual pada pengamatan ke- i
$\underset{\sim}{\varepsilon}$	:	vektor residual ε
E	:	ekspektasi
$\mathbf{\Sigma}$	:	matriks kovarians
$f(x_i)$	:	kurva regresi pada titik x_i
$f(x_{1i}, x_{2i}, \dots, x_{pi})$	:	kurva regresi dengan prediktor x sebanyak p variabel
$f(x_{ji})$	:	kurva regresi dengan prediktor j dan observasi i
f_i	:	kurva regresi pada observasi i
$\underset{\sim}{f}$	:	vektor kurva regresi
f_{K}	:	kurva regresi yang mengandung titik knot K
f_{K_i}	:	kurva regresi yang mengandung titik knot K pada observasi i
$f_{K_{(i)}}$	:	fungsi f_{K_i} yang dihitung tanpa observasi ke- i.
$\hat{f}_K(x)$	:	fungsi penduga dari $f(x)$ yang mengandung titik knot K
$\hat{f}_{K_i}$	:	elemen ke-i dari vektor $\underset{\sim}{f}_K$ berukuran $n \times 1$
$\mathrm{F}_{\mathrm{hitung}}$	:	statistik uji simultan F
$g(\varepsilon_i)$	:	distribusi probabilitas dari ε_i

H_0	:	hipotesis 0
H_1	:	hipotesis 1
i	:	indeks berjalan untuk observasi
$\mathbf{I}$	:	matriks identitas
j	:	indeks berjalan untuk prediktor
k		indeks berjalan untuk titik knot
K_i	:	titik knot ke- i
K_{jr}	:	titik knot pada variabel ke- j sebanyak r
l	:	indeks berjalan untuk orde spline
$L(\underset{\sim}{\beta})$	:	fungsi likelihood
$\ell(\underset{\sim}{\beta})$	:	logaritma dari fungsi likelihood
$L(K)$	:	fungsi *loss*
log	:	logaritma
m	:	derajat polinomial dari *truncated power basis*
MSE	:	*mean squared error* atau rata-rata jumlah kuadrat eror
$\underset{\sim}{\mu}$	:	vektor rata-rata dari $\underset{\sim}{y}$
n	:	banyaknya observasi
N	:	distribusi normal
p	:	banyaknya prediktor
$p(m+r)$	:	jumlah parameter dalam model regresi spline
r	:	banyaknya titik knot
$R(K)$	:	fungsi *risk* dengan titik knot K
$\hat{\mathrm{R}}(\mathrm{K})$	:	kriteria *unbiased risk* dengan titik knot K
R^2	:	koefisien determinasi
s_i	:	elemen diagonal ke-i dari **A(K)**
$\widehat{\mathrm{SE}}(\hat{\beta}_{jl})$	:	estimasi standar eror $\hat{\beta}_{jl}$
$\mathrm{t}_{\mathrm{hitung}}$	:	statistik uji simultan t
tr	:	*trace* atau jumlahan diagonal matriks
$\widehat{var}(\hat{\beta}_{jl})$	:	estimasi varians $\hat{\beta}_{jl}$
x	:	prediktor
x_i	:	prediktor x pada pengamatan ke- i
x_{pi}	:	prediktor x ke- p pada pengamatan ke- i
X_1	:	data untuk prediktor ke-1
X_2	:	data untuk prediktor ke-2
X_3	:	data untuk prediktor ke-3
X_4	:	data untuk prediktor ke-4
X_5	:	data untuk prediktor ke-5

$\mathbf{X}_1$	:	matriks yang memuat variabel x untuk fungsi polinomial pada regresi spline
$\mathbf{X}_2$	:	matriks yang memuat variabel x untuk fungsi *truncated* pada regresi spline
$\mathbf{X}(\mathbf{K})$	:	matriks berukuran $(n \times (1 + p(m + r)))$ dari model yang membentuk penduga f_{K} dan bergantung pada titik knot
$(x_i - \mathrm{K}_r)_+^m$	:	fungsi *truncated* spline dengan titik knot K
y	:	variabel respon
Y	:	data untuk variabel respon
y_i	:	variabel respon y pada pengamatan ke- i
$\underset{\sim}{y}$	:	vektor y
$\underset{\sim}{y}'$	:	*transpose* dari $\underset{\sim}{y}$
$\underset{\sim}{\hat{y}}$	:	estimasi dari $\underset{\sim}{y}$
$\overline{y}$	:	rata-rata y
z_i	:	prediktor z pada pengamatan ke- i
σ^2	:	varians
$\theta(z_i)$	:	fungsi regresi yang tidak diketahui bentuk polanya pada regresi semiparametrik
$\underset{\sim}{0}$	:	vektor berisi bilangan 0
π	:	pi bernilai 3,14159

(halaman ini sengaja dikosongkan)

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Analisis regresi adalah sebuah metode dalam statistika untuk mengetahui hubungan antara variabel respon dan prediktor. Dewasa ini, terdapat tiga pendekatan regresi yang banyak dikembangkan oleh peneliti, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik (Hardle, 1990). Regresi parametrik digunakan apabila bentuk kurva regresinya diketahui, seperti linier, kuadratik, kubik, polinomial, dan lain-lain. Namun, jika bentuk atau pola kurva regresinya tidak diketahui, maka digunakan regresi nonparametrik. Sedangkan regresi semiparametrik merupakan campuran antara regresi parametrik dan regresi nonparametrik, yaitu sebagian bentuk kurva regresi diketahui, dan sebagian lain tidak diketahui.

Pendekatan regresi nonparametrik tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih tinggi. Bentuk estimasi kurva regresi diharapkan menyesuaikan data tanpa dipengaruhi oleh subyektifitas peneliti (Eubank, 1999). Beberapa contoh estimator dalam model regresi nonparametrik yang sering digunakan adalah metode Kernel, Wavelet, Fourier, *Multivariate Adaptif Regression Splines* (MARS) dan Spline. Diantara beberapa model regresi nonparamterik tersebut, metode spline cukup populer di kalangan peneliti karena memiliki beberapa kelebihan, diantaranya adalah memiliki interpretasi visual yang baik, fleksibel, dan dapat menggambarkan perubahan perilaku data pada sub-sub interval tertentu (Eubank, 1999). Budiantara (2009) menyatakan bahwa data yang memiliki pola yang berubah-ubah pada sub-sub interval tertentu sebaiknya dimodelkan menggunakan metode spline.

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 karateristik lokal suatu fungsi atau data (Budiantara, Suryadi, Otok dan Guritno, 2006). Sifat tersegmen ini juga memberikan keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang

tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus (Hardle, 1990). Spline memiliki berbagai pendekatan estimator, diantaranya spline *original*, spline *type* M, spline *relaxed*, spline terbobot, dan lain-lain. Pendekatan spline memiliki suatu basis fungsi, dengan basis fungsi yang biasa dipakai adalah spline *truncated* dan B-Spline (Lyche and Morken, 2004, dalam Budiantara dkk., 2006).

Beberapa hal yang perlu diperhatikan dalam membentuk model pada regresi spline adalah menentukan orde model, banyaknya titik knot dan lokasi titik knot tersebut (Montoya, Ulloa dan Miller, 2014). Orde model dapat diketahui dari bentuk pola antar variabel. Yang tidak kalah pentingnya adalah menentukan banyak titik knot dan lokasi dari tiap-tiap knot karena akan berpengaruh pada bentuk kurva regresi. Ada beberapa metode yang dapat digunakan untuk memilih titik knot optimal dalam estimator spline, antara lain metode *Cross Validation* (CV) (Craven dan Wahba, 1979), *Unbiased Risk* (UBR) (Wahba, 1990; Wang, 1998), *Generalized Maximum Likelihood* (GML) (Eubank, 1999) dan *Generalized Cross Validation* (GCV) (Wahba, 1990).

Salah satu metode yang kerap digunakan dalam pemilihan titik knot optimal adalah GCV. Metode GCV memberikan beberapa kelebihan dibandingkan metode lainnya, diantaranya memiliki sifat optimal asimptotik, invarian terhadap transformasi dan dalam perhitungannya varians populasi tidak perlu diketahui (Wahba, 1990). CV adalah suatu metode dimana data digunakan kembali dengan cara tertentu yang mensimulasikan apa yang akan dilakukan jika pengamatan berulang tersedia (Eubank, 1999). Kelebihan metode CV adalah nilai estimasi pada CV dihitung dari regresi lokal yang menghilangkan suatu observasi (Fox, 2008). Metode GCV dan CV merupakan metode yang terkenal karena sifat optimalnya (Wahba, 1990 dalam Opsomer, Wang dan Yang, 2001). Metode GCV merupakan metode CV yang terboboti, yaitu dengan memberikan bobot yang sama pada setiap pengamatan, sedangkan CV merupakan regresi terboboti dengan bobot yang berbeda-beda pada masing-masing pengamatan sesuai dengan kontribusinya, sehingga metode CV diharapkan memberikan hasil yang lebih akurat dalam menentukan estimatornya.

UBR merupakan suatu metode pemilihan parameter penghalus yang mensyaratkan nilai estimasi dari varian eror diketahui (Wahba, 1990; Wang, 2011). Metode UBR baik digunakan untuk memilih parameter pemulus untuk estimasi spline dengan data *non-Gaussian* (Gu, 1992; Wahba, Wang, Gu, Klein dan Klein, 1995 dalam Wang, 1998). Eksperimen numerik yang dilakukan Craven dan Wahba (1979) menunjukkan bahwa estimasi GCV dan UBR pada dasarnya sama dalam menghasilkan akurasi eksperimen ketika varian yang sama digunakan untuk menghasilkan data eksperimen (Wahba, 1990). Sementara itu, metode GML merupakan pengembangan estimasi berbasis *likelihood* untuk parameter *smoothing* dengan memanfaatkan hubungan antara model spline *smoothing* dan model Bayesian. Berdasarkan simulasi yang dilakukan oleh Kohn, Ansley dan Tharm (1991), GML kompetitif, bahkan superior dibandingkan GCV ketika menyeleksi penghalus. Metode GML juga cukup stabil dan efisien untuk jumlah sampel kecil sampai menengah (Opsomer dkk., 2001). Namun, metode GML cenderung *undersmooth* ketika menyeleksi penghalus untuk data dari model frekuentis (Wahba, 1990). Asumsi normalitas juga diperlukan dalam derivasi kriteria GML, sementara kriteria GCV dan UBR tidak memerlukan asumsi distribusi (Wang, 2011). Kekurangan metode GML, GCV dan UBR yaitu metode tersebut dikembangkan di bawah asumsi pengamatan independen, sehingga cenderung mengabaikan parameter penghalus ketika data berkorelasi (Wang, 1998). Wahba (1990) juga menyatakan bahwa penggunaan GCV adalah asimptotik, sehingga hasil yang baik tidak dapat dicapai ketika ukuran sampel sangat kecil.

Penelitian ini menggunakan regresi nonparametrik spline *truncated* untuk pemodelan pada data tingkat pengangguran terbuka di Jawa Tengah. Masalah pengangguran selalu menjadi permasalahan krusial yang dihadapi oleh semua negara. Tingginya jumlah pengangguran dikarenakan jumlah lapangan kerja yang tidak lagi mampu mengimbangi banyaknya jumlah angkatan kerja yang terus bertambah tiap tahunnya. Tingkat pengangguran suatu negara dapat memberikan gambaran kasar apakah perekonomian negara tersebut berkembang atau justru mengalami kemunduran. Apabila pengangguran di suatu negara sangat buruk, kekacauan politik dan sosial selalu berlaku dan menimbulkan efek yang buruk

bagi kesejahteraan masyarakat dan prospek pembangunan ekonomi dalam jangka panjang (Sukirno, 2004).

Jawa Tengah merupakan salah satu provinsi yang mulai diminati oleh para investor untuk pendirian industri, baik investor dalam negeri maupun asing. Pada semester pertama tahun 2017, realisasi investasi di Jawa Tengah mencapai 20,44 triliun rupiah atau 49% dari yang ditargetkan sampai akhir tahun 2017, yaitu 41,70% (Sujianto, 2017). Namun, Provinsi Jawa Tengah masih mengalami masalah pengangguran yang cukup memprihatinkan. Tingkat pengangguran terbuka di Provinsi Jawa Tengah menduduki peringkat 19 tertinggi dalam skala nasional. Jumlah pengangguran di Jawa Tengah pada Februari 2017 adalah sebesar 0,76 juta orang dengan tingkat pengangguran terbuka sebesar 4,15% (BPS, 2017). Angka ini masih lebih tinggi jika dibandingkan dengan provinsi lainnya di Pulau Jawa, seperti Jawa Timur dan Yogyakarta. Tingginya angka pengangguran di Jawa Tengah perlu mendapat perhatian lebih dari pemerintah, mengingat Provinsi Jawa Tengah memiliki potensi tinggi untuk mengembangkan sumber daya manusianya.

Beberapa penelitian mengenai tingkat pengangguran terbuka dan faktor-faktor yang mempengaruhinya pernah dilakukan oleh beberapa peneliti. Penelitian Prahutama (2013) yang menggunakan pendekatan nonparametrik deret fourier dan penelitian Sari dan Budiantara (2012) yang menggunakan pendekatan regresi spline multivariabel memakai variabel Angka Partisipasi Kasar (APK) dan upah minimum sebagai variabel yang mempengaruhi pengangguran terbuka di Jawa Timur. Astuti, Ratnasari dan Wibowo (2017) dalam penelitiannya mengenai regresi data panel menyimpulkan bahwa variabel upah minimum dan Tingkat Partisipasi Angkatan Kerja (TPAK) berpengaruh signifikan terhadap pengangguran di Jawa Timur. Sedangkan Biro Analisa Anggaran dan Pelaksanaan APBN menyatakan adanya keterkaitan antara inflasi dan pertumbuhan ekonomi terhadap tingkat pengangguran. Dari referensi dan penelitian-penelitian tersebut mendasari penulis memilih variabel inflasi, laju pertumbuhan ekonomi, APK, upah minimum, dan TPAK sebagai prediktor yang mempengaruhi variabel tingkat pengangguran terbuka.

Pemodelan tingkat pengangguran terbuka pada penelitian ini menggunakan pengamatan kota/kabupaten di Provinsi Jawa Tengah. Fungsi kriteria GCV memiliki pembagi yang bersifat global, sehingga baik digunakan untuk data yang memiliki variansi populasi yang sama. Namun, untuk pengamatan kota/kabupaten cenderung memiliki variansi populasi yang berbeda, sehingga akan lebih baik jika menggunakan metode CV yang memiliki bobot yang berbeda pada setiap pengamatan. Metode lain yang dapat digunakan untuk mengakomodasi permasalahan tersebut adalah UBR, sehingga penelitian ini akan berfokus pada pemilihan titik knot menggunakan metode CV dan UBR. Pengkajian mengenai CV dan UBR telah dilakukan oleh Eubank (1999), Wahba (1990) dan Wang (1998) meskipun terbatas pada pengkajian *smoothing splines* pada data univariabel. Penelitian ini mengkaji CV dan UBR untuk pemilihan titik knot pada data multivariabel yang diharapkan akan menjadi suatu kajian yang penting bagi pengembangan ilmu statistika.

1.2 Rumusan Masalah

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

1. Bagaimana penurunan kriteria UBR sehingga ekspektasi UBR sama dengan fungsi *risk* dan penurunan fungsi CV pada regresi nonparametrik multivariabel spline *truncated*?
2. Bagaimana perbandingan kinerja metode UBR dan CV dalam menduga titik knot optimal pada regresi nonparametrik multivariabel spline *truncated* pada data simulasi?
3. Bagaimana aplikasi pemilihan titik knot optimal menggunakan metode UBR dan CV dalam regresi nonparametrik multivariabel spline *truncated* pada data tingkat pengangguran terbuka di Provinsi Jawa Tengah?

1.3 Tujuan

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengkaji fungsi UBR dan CV pada regresi nonparametrik multivariabel spline *truncated.*
2. Melakukan studi simulasi pada model spline *truncated* multivariabel untuk membandingkan kinerja metode UBR dan CV dalam mengestimasi titik knot optimal.
3. Membandingkan metode UBR dan CV untuk memilih titik knot optimal pada regresi nonparametrik multivariabel spline *truncated* menggunakan data tingkat pengangguran terbuka di Provinsi Jawa Tengah.

1.4 Manfaat

Penelitian ini diharapkan dapat memberi manfaat sebagai berikut:

1. Menambah wawasan keilmuan dan pengetahuan tentang metode UBR dan CV.
2. Memberikan informasi kepada pemerintah untuk mengetahui variabel-variabel yang mempengaruhi tingkat pengangguran terbuka di Provinsi Jawa Tengah.
3. Hasil penelitian diharapkan menjadi bahan masukan dan acuan untuk penelitian-penelitian selanjutnya.

1.5 Batasan Masalah

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

1. Model regresi nonparametrik yang digunakan adalah spline *truncated* linier.
2. Dalam aplikasi digunakan spline *truncated* dengan satu, dua dan tiga kombinasi knot.

BAB 2
TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi adalah studi mengenai ketergantungan variabel dependen (respon) dengan satu atau lebih variabel independen (prediktor) dengan tujuan untuk mengestimasi dan/atau memprediksi rata-rata populasi atau nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui (Gujarati, 2003). Hubungan antara respon dan prediktor untuk *n* buah pengamatan, dengan pengamatan $(x_i, y_i),\ i = 1,2, \dots, n$ adalah:

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

dimana y_i adalah respon y pada pengamatan ke-i, x_i adalah prediktor x pada pengamatan ke-i, ε_i adalah eror atau residual pada pengamatan ke-i yang merupakan variabel random independen dengan mean 0 dan varians konstan σ^2, $f(x_i)$ adalah kurva regresi pada titik x_i (Eubank, 1999).

Jika bentuk kurva regresi $f(x_i)$ diketahui, maka regresi tersebut merupakan regresi parametrik. Kurva regresi dalam regresi parametrik dapat menghasilkan pola yang linier maupun non-linier. Regresi nonparametrik digunakan jika bentuk kurva regresi $f(x_i)$ tidak diketahui. Regresi nonparametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu. Regresi nonparametrik memiliki fleksibilitas yang tinggi karena bentuk estimasi kurva regresinya menyesuaikan datanya tanpa dipengaruhi oleh faktor subyektifitas peneliti (Eubank, 1999). Asumsi dasar dari regresi nonparametrik yakni adanya penghalus yang menghubungkan respon *y* terhadap satu atau lebih prediktor *x*. Beberapa metode regresi nonparametrik yang telah dikembangkan antara lain adalah regresi kernel, polinomial lokal, wavelet, fourier dan spline.

Sementara itu, regresi semiparametrik merupakan gabungan dari komponen parametrik dan komponen nonparametrik. Model regresi semiparametrik yang memiliki pengamatan $(x_i, z_i, y_i),\ i = 1,2, \dots, n$ adalah:

$$y_i = f(x_i) + \theta(z_i) + \varepsilon_i, \quad i = 1,2, \dots, n \tag{2.2}$$

dimana y_i adalah respon, x_i dan z_i adalah prediktor, dan ε_i adalah eror random independen dengan mean 0 dan variansi konstan σ^2, $f(x_i)$ adalah fungsi regresi yang diketahui bentuk polanya sedangkan $\theta(z_i)$ adalah fungsi regresi yang tidak diketahui bentuk polanya.

2.2 Regresi Spline *Truncated*

Beberapa penulis ternama seperti Hardle (1990) dan Wahba (1990) menyarankan penggunaan regresi spline sebagai alternatif pendekatan non parametrik. Spline merupakan potongan (*truncated*) polinomial tersegmen yang kontinu, sehingga mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus (Hardle, 1990). Spline dalam regresi nonparametrik mempunyai sifat fleksibel dan mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan (Eubank, 1999; Budiantara, 2005). Hardle (1990) mendefinisikan titik knot sebagai perpaduan bersama yang menunjukkan pola perilaku fungsi spline pada selang yang berbeda. Dalam spline digunakan *truncated power basis* dengan *r* knot, misalkan $K_1, K_2, \dots, K_r$ yaitu:

$$\{1, x, x^2, \dots, x^m \ (x-K_1)_+^m, (x-K_2)_+^m, \dots, (x-K_r)_+^m\},$$

dimana m menunjukkan derajat polinomial dari *truncated power basis*. *Truncated power basis* merupakan potongan fungsi spline orde m yang tersegmen kontinyu sehingga membantu dalam mencari estimasi yang optimal melalui titik knot yang membentuknya. Derajat m = 1, 2, 3 secara umum merupakan *truncated power basis* linier, kuadratik dan kubik.

Menurut Wu dan Zang (2006), secara umum fungsi *spline truncated* derajat *m* dengan titik knot $(K_1, K_2, ..., K_r)$ disajikan dalam fungsi $f(x)$ dengan persamaan berikut:

$$f(x) = \sum_{l=0}^{m} \beta_l x^l + \sum_{k=1}^{r} \beta_{k+m} (x - K_k)_+^m \qquad (2.3)$$

dengan fungsi *truncated*,

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

Persamaan tersebut memiliki sejumlah r titik knot $a < K_1 < K_2 < ... < K_r < b$, dimana a merupakan nilai minimum dari x dan b nilai maksimum dari x. Apabila diberikan model nonparametrik :

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

dimana ε_i adalah residual random independen dengan mean nol dan varians σ^2. Maka, bentuk umum regresi spline keluarga polinomial *truncated* orde ke-m dengan satu prediktor adalah

$$y_i = \sum_{l=0}^{m} \beta_l x_i^{\,l} + \sum_{k=1}^{r} \beta_{k+m} (x_i - K_k)_+^m + \varepsilon_i, \quad i = 1,2, \dots, n . \tag{2.4}$$

Dengan menggunakan data amatan sebanyak n, bentuk persamaan (2.4) dapat ditulis ke dalam matriks :

$$\underset{\sim}{y} = \mathbf{X}_1 \underset{\sim}{\delta}_1 + \mathbf{X}_2 \underset{\sim}{\delta}_2 + \underset{\sim}{\varepsilon} \tag{2.5}$$

dimana,

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{nx1} ; \mathbf{X}_1 = \begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^m \end{bmatrix}_{n \times m} ; \underset{\sim}{\delta}_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}_{m \times 1} ; \underset{\sim}{\delta}_2 = \begin{bmatrix} \beta_{m+1} \\ \beta_{m+2} \\ \vdots \\ \beta_{m+r} \end{bmatrix}_{r \times 1}$$

$$\mathbf{X}_2 = \begin{bmatrix} (x_1-K_1)_+^m & (x_1-K_2)_+^m & \cdots & (x_1-K_r)_+^m \\ (x_2-K_1)_+^m & (x_2-K_2)_+^m & \cdots & (x_2-K_r)_+^m \\ \vdots & \vdots & \ddots & \vdots \\ (x_n-K_1)_+^m & (x_n-K_2)_+^m & \cdots & (x_n-K_r)_+^m \end{bmatrix}_{n \times r} ; \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

Bentuk matriks dari persamaan regresi spline (2.5) dapat disederhanakan menjadi

$$\underset{\sim}{y} = \mathbf{X} \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{2.6}$$

dengan $\mathbf{X} = \left[\mathbf{X_1} \vdots \mathbf{X_2} \right]$ dan $\underset{\sim}{\beta} = \begin{bmatrix} \underset{\sim}{\delta}_1 \\ \cdots \\ \underset{\sim}{\delta}_2 \end{bmatrix}$

2.3 Regresi Spline *Truncated* Multivariabel

Dalam analisis regresi nonparametrik spline, jika terdapat satu respon dan satu prediktor maka regresi tersebut disebut regresi nonparametrik spline univariabel. Sedangkan apabila terdapat satu respon dengan lebih dari satu prediktor, maka regresi tersebut disebut regresi nonparametrik spline multivariabel (Budiantara, 2004). Misalkan diberikan data $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ dimana hubungan antara $(x_{1i}, x_{2i}, \dots, x_{pi})$ dan y_i diasumsikan mengikuti model regresi nonparametrik, model regresi nonparametrik spline *truncated* multivariabel adalah sebagai berikut:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i \, , i = 1,2, \dots, n \tag{2.7}$$

dimana f adalah kurva regresi yang tidak diketahui bentuknya. Apabila kurva regresi f merupakan model aditif dan dihampiri dengan fungsi spline, maka diperoleh model regresi sebagai berikut:

$$y_i = \sum_{j=1}^{p} f\left(x_{ji}\right) + \varepsilon_i \, , i = 1,2, \dots, n \tag{2.8}$$

dimana

$$f\left(x_{ji}\right) = \beta_{j0} + \sum_{l=1}^{m} \beta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \beta_{j\,(m+k)} (x_{ji} - K_{jk})_{+}^{m} \, , j = 1,2, \dots, p$$

dengan

$$(x_{ji} - K_{jk})_{+}^{m} = \begin{cases} (x_{ji} - K_{jk})^{m} & , x_{ji} \geq K_{jk} \\ 0 & , x_{ji} < K_{jk} \end{cases}$$

dimana persamaan tersebut memiliki sejumlah *r* titik knot yaitu $K_{j1}, \; K_{j2}, \dots, K_{jr}$.

2.4 Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola pada interval yang berlainan. Bentuk estimator spline sangat dipengaruhi oleh nilai parameter penghalus (Budiantara, 2005). Bentuk estimator spline juga dipengaruhi oleh lokasi dan banyaknya titik-titik knot. Oleh karena itu, pemilihan titik knot optimal mutlak diperlukan untuk memperoleh estimator spline yang sesuai dengan data. Ada beberapa metode yang dapat digunakan untuk memilih parameter penghalus optimal dalam estimator spline, antara lain metode *Cross Validation* (Craven dan Wahba, 1979), *Unbiased Risk* (Wahba,

1990; Wang,1998), *Generalized Maximum Likelihood* (Eubank, 1999) dan *Generalized Cross Validation* (Wahba, 1990). Penelitian ini menggunakan pemilihan titik knot UBR dan CV. Pemilihan titik knot dengan metode UBR didapatkan dengan meminimalkan kriteria UBR, dimana kriteria UBR diberikan oleh fungsi berikut:

$$\hat{\mathrm{R}}(\mathrm{K}) = \frac{1}{n} \| \mathbf{(I - A(K))} \underset{\sim}{y} \|^2 - \frac{\sigma^2}{n} \mathrm{tr}\mathbf{(I - A'(K))(I - A(K))} + \frac{\sigma^2}{n} \mathrm{tr}\mathbf{(A'(K))(A(K))}. \quad (2.9)$$

Pemilihan titik knot dengan metode CV didapatkan dengan meminimalkan fungsi CV. Fungsi CV untuk pemilihan titik knot optimal dapat ditunjukkan pada persamaan berikut:

$$\mathrm{CV(K)} = \frac{1}{n} \sum_{i=1}^{n} (y_i - f_{K(i)})^2 . \quad (2.10)$$

2.5 Pengujian Parameter Model

Uji parameter model dilakukan untuk melihat apakah prediktor berpengaruh pada respon. Dalam penelitian ini, uji parameter model dilakukan setelah didapatkan model regresi nonparamterik spline *truncated* dengan titik knot optimal berdasarkan kriteria nilai UBR atau CV minimum. Pengujian parameter model dibagi menjadi uji serentak dan uji parsial. Pengujian parameter model secara serentak merupakan uji signifikansi seluruh parameter dalam model secara keseluruhan. Hipotesis yang dipakai adalah sebagai berikut:

H_0: $\beta_{11} = \beta_{12} = \cdots = \beta_{p(m+r)} = 0$

H_1: minimal ada satu $\beta_{jl} \neq 0, \quad j = 1,2,\ldots,p; \quad l = 1,2,\ldots,m+r$

Nilai $\mathrm{F_{hitung}}$ didapatkan dengan rumus berikut:

$$\mathrm{F_{hitung}} = \frac{\mathrm{MS_{regresi}}}{\mathrm{MS_{residual}}} \quad (2.11)$$

dimana daerah penolakannya adalah tolak H_0 jika

$$\mathrm{F_{hitung}} > \mathrm{F}_{\propto;p(m+r);n-1-p(m+r)}.$$

Nilai $p(m+r)$ adalah jumlah parameter dalam model regresi sedangkan n merupakan jumlah observasi. Perhitungan nilai statistik uji F didapatkan dari Analisis Ragam (ANOVA) pada Tabel 2.1.

Tabel 2.1 Analisis Ragam (ANOVA)

Sumber Variasi	**Derajat Bebas (df)**	**Jumlah Kuadrat (SS)**	**Rataan Kuadrat (MS)**	$\mathbf{F_{hitung}}$
Regresi	$p(m+r)$	$\hat{\underset{\sim}{\beta}}' \mathbf{X}' \underset{\sim}{y} - n\bar{y}^2$	$\frac{\mathrm{SS_{regresi}}}{\mathrm{df_{regresi}}}$	
Residual	$n-1-p(m+r)$	$\underset{\sim}{y}' \underset{\sim}{y} - \hat{\underset{\sim}{\beta}}' \mathbf{X}' \underset{\sim}{y}$	$\frac{\mathrm{SS_{residual}}}{\mathrm{df_{residual}}}$	$\frac{\mathrm{MS_{regresi}}}{\mathrm{MS_{residual}}}$
Total	$n-1$	$\underset{\sim}{y}' \underset{\sim}{y} - n\bar{y}^2$		

Pengujian parameter model secara parsial merupakan uji untuk melihat apakah masing-masing prediktor berpengaruh secara signifikan pada respon. Hipotesis yang dipakai adalah sebagai berikut:

$H_0 : \beta_{jl} = 0$

$H_1 : \beta_{jl} \neq 0, \quad j = 1,2,\dots,p; \quad l = 1,2,\dots,m+r$

Statistik uji:

$$\mathrm{t_{hitung}} = \frac{\hat{\beta}_{jl}}{\widehat{\mathrm{SE}}(\hat{\beta}_{jl})} \tag{2.12}$$

dengan

$$\widehat{\mathrm{SE}}(\hat{\beta}_{jl}) = \sqrt{\widehat{var}(\hat{\beta}_{jl})} \tag{2.13}$$

dimana $\widehat{var}(\hat{\beta}_{jl})$ merupakan elemen diagonal $(\mathbf{X^T X})^{-1} MSE$. Daerah penolakan dari persamaan (2.12) yaitu:

$$\left|\mathrm{t_{hitung}}\right| \geq \mathrm{t}_{(\frac{\propto}{2}, n-1-p(m+r))} . \tag{2.14}$$

2.6 *Mean Squared Error* (MSE) dan Koefisien Determinasi

MSE atau rata-rata jumlah kuadrat eror merupakan nilai yang menunjukkan jarak atau selisih antara nilai estimasi dengan nilai sebenarnya. MSE merupakan suatu kriteria kebaikan suatu model yang memiliki nilai non negatif. Semakin kecil nilai MSE yang dihasilkan suatu model, maka semakin baik model tersebut. Rumus untuk menghitung nilai MSE pada model regresi spline *truncated* multivariabel adalah sebagai berikut:

$$\text{MSE} = \frac{\underset{\sim}{y}'\underset{\sim}{y} - \hat{\underset{\sim}{\beta}}'\mathbf{X}'\underset{\sim}{y}}{n-1-p(m+r)} \qquad (2.15)$$

Koefisien determinasi (R^2) digunakan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen. Nilai R^2 berada antara nol dan satu. Rumus untuk menghitung R^2 adalah sebagai berikut:

$$R^2 = \frac{SSR}{SST} \qquad (2.16)$$

dimana

$$SSR = \hat{\underset{\sim}{\beta}}'\mathbf{X}'\underset{\sim}{y} - n\bar{y}^2 \ , \ SST = \underset{\sim}{y}'\underset{\sim}{y} - n\bar{y}^2$$

Semakin tinggi nilai R^2 yang dihasilkan suatu model, maka semakin baik prediktor dalam model tersebut dalam menjelaskan variabilitas respon.

2.7 Nilai Ekspektasi dari Bentuk Kuadratik

Sub bab ini menjelaskan mengenai ekspektasi dari bentuk kuadratik $\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}$ yang dikutip dari Rencher dan Scaalje (2007).

Teorema 1. Jika $\underset{\sim}{y}$ adalah vektor random dengan rata-rata $\underset{\sim}{\mu}$ matriks kovarians $\mathbf{\Sigma}$ dan apabila *A* adalah matriks simetrik yang berisi konstanta, maka

$$E(\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}) = \text{tr}(\mathbf{A\Sigma}) + \underset{\sim}{\mu}'\mathbf{A}\underset{\sim}{\mu}.$$

Persamaan tersebut dapat dibuktikan sebagai berikut:

Apabila

$$\mathbf{\Sigma} = E[(\underset{\sim}{y} - \underset{\sim}{\mu})(\underset{\sim}{y} - \underset{\sim}{\mu})']$$

$$= E(\underset{\sim}{y}\underset{\sim}{y}') - \underset{\sim}{\mu}\underset{\sim}{\mu}'.$$

Sehingga dapat ditulis menjadi:

$$E(\underset{\sim}{y}\underset{\sim}{y}') = \mathbf{\Sigma} + \underset{\sim}{\mu}\underset{\sim}{\mu}'.$$

Karena $\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}$ merupakan skalar, maka nilainya sama dengan *trace*. Sehingga didapatkan:

$$E(\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}) = E[\text{tr}(\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y})]$$

$$= E[\text{tr}(\mathbf{A}\underset{\sim}{y}\underset{\sim}{y}')]$$

$$=\text{tr}[\text{E}(\mathbf{A}\underset{\sim}{y}\underset{\sim}{y}')]$$

$$=\text{tr}[\mathbf{A}\text{E}(\underset{\sim}{y}\underset{\sim}{y}')]$$

$$=\text{tr}[\mathbf{A}(\mathbf{\Sigma}+\underset{\sim}{\mu}\underset{\sim}{\mu}')]$$

$$=\text{tr}[\mathbf{A}\mathbf{\Sigma}+\mathbf{A}\underset{\sim}{\mu}\underset{\sim}{\mu}']$$

$$=\text{tr}(\mathbf{A}\mathbf{\Sigma})+\text{tr}(\underset{\sim}{\mu}'\mathbf{A}\underset{\sim}{\mu})$$

$$=\text{tr}(\mathbf{A}\mathbf{\Sigma})+\underset{\sim}{\mu}'\mathbf{A}\underset{\sim}{\mu}.$$

Perlu diperhatikan bahwa $\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}$ bukan merupakan fungsi linier dari $\underset{\sim}{y}$ sehingga

$$\text{E}(\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}) \neq \text{E}(\underset{\sim}{y}')\mathbf{A}\text{E}(\underset{\sim}{y}).$$

2.8 Tingkat Pengangguran Terbuka (TPT)

Badan Pusat Statistik (2007) menyatakan bahwa tenaga kerja adalah modal penting bagi bergeraknya roda pembangunan di suatu negara. Oleh karena itu, untuk mengetahui keberhasilan pembangunan di suatu negara dibutuhkan informasi mengenai statistik ketenagakerjaan, seperti Tingkat Pengangguran Terbuka (TPT). TPT merupakan presentase antara jumlah pengangguran terhadap jumlah angkatan kerja. Penduduk yang masuk dalam kategori angkatan kerja adalah penduduk usia kerja (15 tahun atau lebih) yang bekerja, atau memiliki pekerjaan tapi sementara tidak bekerja dan pengangguran. Penduduk bukan angkatan kerja adalah penduduk usia kerja (15 tahun atau lebih) yang masih sekolah, mengurus rumah tangga atau melaksanakan kegiatan lain selain kegiatan pribadi.

Masalah pengangguran selalu menjadi permasalahan krusial yang dihadapi oleh semua negara. Apabila pengangguran di suatu negara sangat buruk, kekacauan politik dan sosial terjadi dan menimbulkan efek yang buruk bagi kesejahteraan masyarakat dan prospek pembangunan ekonomi dalam jangka panjang (Sukirno, 2004). Semakin tinggi angka pengangguran terbuka, semakin besar potensi kerawanan sosial seperti kriminalitas. Sebaliknya, semakin rendah angka pengangguran terbuka, semakin stabil kondisi sosial dalam masyarakat.

Karena itulah TPT kerap dijadikan indikator sebagai tolok ukur keberhasilan pembangunan.

2.9 Faktor-faktor yang Mempengaruhi TPT

Ada beberapa faktor yang diduga mempengaruhi tingkat pengangguran terbuka. Haug dan King (2014) dan Bhattarai (2016) menjelaskan adanya hubungan inflasi dan pengangguran. Bank Indonesia mendefinisikan inflasi sebagai meningkatnya harga secara umum dan terus menerus. Salah satu indikator yang sering digunakan dalam perhitungan inflasi adalah Indeks Harga Konsumen (IHK). Bank Indonesia menjelaskan pengendalian inflasi penting dilakukan, karena inflasi yang tinggi dan tidak stabil memberi dampak negatif pada kondisi sosial ekonomi masyarakat, seperti menyebabkan pendapatan riil masyarakat terus menurun, menciptakan ketidakpastian (*uncertainty*) bagi pelaku ekonomi dalam mengambil keputusan, dan juga memberikan tekanan pada nilai rupiah jika inflasi dalam negeri lebih tinggi daripada inflasi di negara tetangga.

Muslim (2014) dan Biro Analisa Anggaran dan Pelaksanaan APBN menyatakan adanya keterkaitan antara pertumbuhan ekonomi terhadap tingkat pengangguran. Menurut Simon Kuznets (dalam Jhingan, 2000), pertumbuhan ekonomi merupakan kenaikan jangka panjang kemampuan suatu negara untuk menyediakan semakin banyak jenis barang-barang ekonomi bagi penduduknya. Secara umum, pertumbuhan ekonomi didefinisikan sebagai peningkatan suatu perekonomian dalam memproduksi barang dan jasa.

Snieska, Valodkiene, Daunoriene dan Draksaite (2015) dan Lavrinovicha, Lavrinenko dan Treinovskis (2015) menyatakan adanya hubungan antara pendidikan dan tingkat ketenagakerjaan, begitu juga dengan penelitian Prahutama (2013) dan Sari dan Budiantara (2012) yang menggunakan variabel Angka Partisipasi Kasar (APK) untuk pemodelan TPT di Jawa Timur. BPS mendefinisikan Angka Partisipasi Kasar (APK) sebagai proporsi anak sekolah pada suatu jenjang tertentu dalam kelompok usia yang sesuai dengan jenjang pendidikan tersebut. Indikator APK digunakan untuk menunjukkan tingkat partisipasi penduduk secara umum pada suatu tingkat pendidikan. Nilai APK yang tinggi menunjukkan tingginya tingkat partisipasi sekolah, tanpa memperhatikan

ketepatan usia sekolah pada jenjang pendidikannya. Pada penelitian ini akan digunakan indikator APK untuk tingkat Sekolah Menengah Pertama (SMP).

Slonimczyk dan Skott (2012), Gorry (2013), Astuti (2017), Prahutama (2013) dan Sari dan Budiantara (2012) dalam penelitiannya menyatakan bahwa ada hubungan antara upah minimum dan tingkat pengangguran. Tujuan penentuan upah minimum antara lain menonjolkan arti dan peranan tenaga kerja (buruh) sebagai sub sistem yang kreatif dalam suatu sistem kerja (Kartasapoetra, 1992). Selain itu, upah minimum berguna untuk melindungi kelompok kerja dari adanya sistem pengupahan yang sangat rendah dan yang keadaannya secara material kurang memuaskan.

Astuti dkk. (2017) dalam penelitiannya menyatakan Tingkat Partisipasi Angkatan Kerja (TPAK) berpengaruh signifikan terhadap tingkat pengangguran. Feng, Hu dan Moffit (2017) menyatakan adanya hubungan antara TPAK dan TPT di Cina. Hubungan jangka panjang antara TPAK dan TPT terdapat di negara Swedia (Osterholm, 2010). TPAK adalah keterlibatan penduduk dalam kegiatan ekonomi diukur dengan porsi penduduk yang masuk dalam pasar kerja (bekerja atau mencari kerja). TPAK merupakan suatu ukuran proporsi penduduk usia kerja yang terlibat secara aktif dalam pasar tenaga kerja, baik yang bekerja maupun sedang mencari pekerjaan (BPS, 2007). Semakin tinggi TPAK mengindikasikan semakin besar pasokan tenaga kerja (*labor supply*) yang tersedia untuk memproduksi barang dan jasa dalam suatu perekonomian.

BAB 3
METODOLOGI PENELITIAN

3.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang bersumber dari Badan Pusat Statistik Provinsi Jawa Tengah tahun 2015. Variabel yang digunakan pada penelitian ini terdiri dari dua jenis, yaitu respon (Y) dan prediktor (X). Unit observasi meliputi 35 kabupaten/kota di Provinsi Jawa Tengah pada tahun 2015. Variabel-variabel penelitian yang digunakan disajikan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

	Variabel Penelitian	**Skala Data**
	(1)	(2)
Y	Tingkat Pengangguran Terbuka	Rasio
X_1	Inflasi	Rasio
X_2	Laju Pertumbuhan Ekonomi	Rasio
X_3	Angka Partisipasi Kasar SMP	Rasio
X_4	Upah Minimum Kota/Kabupaten	Rasio
X_5	Tingkat Partisipasi Angkatan Kerja	Rasio

Adapun definisi operasional dari variabel-variabel yang digunakan dalam penelitian ini disajikan pada Tabel 3.2.

Tabel 3.2 Definisi Operasional Variabel Penelitian

Variabel Penelitian	**Definisi Operasional**
(1)	(2)
Tingkat Pengangguran Terbuka	Tingkat Pengangguran Terbuka (TPT) merupakan presentase antara jumlah pengangguran terhadap jumlah angkatan kerja. Penduduk yang masuk dalam kategori angkatan kerja adalah penduduk usia kerja (15 tahun atau lebih) yang bekerja, atau memiliki pekerjaan tapi sementara tidak bekerja dan pengangguran.
Inflasi	Inflasi adalah meningkatnya harga secara umum dan terus menerus. Kenaikan harga dari satu atau dua barang saja tidak dapat disebut inflasi kecuali bila kenaikan itu meluas (atau mengakibatkan kenaikan harga) pada barang lainnya. Salah satu indikator yang sering digunakan dalam perhitungan inflasi adalah Indeks Harga Konsumen (IHK).

Tabel 3.2 Definisi Operasional Variabel Penelitian (Lanjutan)

Variabel Penelitian	Definisi Operasional
(1)	(2)
Laju Pertumbuhan Ekonomi	Secara umum, pertumbuhan ekonomi didefinisikan sebagai peningkatan suatu perekonomian dalam memproduksi barang dan jasa. Laju pertumbuhan ekonomi diukur melalui indikator perkembangan Produk Domestik Bruto (PDB) dari tahun ke tahun.
Angka Partisipasi Kasar SMP	Angka Partisipasi Kasar (APK) merupakan proporsi anak sekolah pada suatu jenjang tertentu dalam kelompok usia yang sesuai dengan jenjang pendidikan tersebut. Indikator APK digunakan untuk menunjukkan tingkat partisipasi penduduk secara umum pada suatu tingkat pendidikan.
Upah Minimum Kota/ Kab.	Upah minimum adalah upah bulanan terendah yang terdiri atas upah pokok termasuk tunjangan tetap yang ditetapkan oleh gubernur sebagai jaring pengaman.
Tingkat Partisipasi Angkatan Kerja	Tingkat partisipasi angkatan kerja yaitu keterlibatan penduduk dalam kegiatan ekonomi diukur dengan porsi penduduk yang masuk dalam pasar kerja (bekerja atau mencari kerja)

3.2 Struktur Data Penelitian

Unit observasi yang digunakan dalam penelitian ini adalah sebanyak 35 kabupaten/kota di Provinsi Jawa Tengah dengan empat buah prediktor. Sehingga, struktur data penelitian yang digunakan adalah sebagai berikut :

Tabel 3.3 Struktur Data Penelitian

Kabupaten/ Kota	$\boldsymbol{Y}$	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$	$\boldsymbol{X_5}$
1	Y_1	X_{11}	X_{21}	X_{31}	X_{41}	X_{51}
2	Y_2	X_{12}	X_{22}	X_{32}	X_{42}	X_{52}
3	Y_3	X_{13}	X_{23}	X_{33}	X_{43}	X_{53}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	Y_{35}	$X_{1(35)}$	$X_{2(35)}$	$X_{3(35)}$	$X_{4(35)}$	$X_{5(35)}$

3.3 Langkah-Langkah Penelitian

Untuk menjawab tujuan penelitian maka secara umum langkah-langkah yang akan dilakukan dalam penelitian ini yaitu sebagai berikut :

3.3.1 Kajian Teoritis

Tujuan pertama dari penelitian ini yaitu mengkaji pemilihan titik knot optimal menggunakan metode *Unbiased Risk* (UBR) dan *Cross Validation* (CV) pada regresi nonparametrik multivariabel spline *truncated.* Tahapan penelitian yang dilakukan untuk mencapai tujuan tersebut adalah sebagai berikut :

a) Mengkaji penurunan metode UBR dari fungsi *loss* dan *risk*

1. Definisikan model regresi

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

2. Definisikan fungsi *loss*

$$L(K) = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - f_{K_i} \right)^2$$

3. Definisikan fungsi *risk*, yang merupakan nilai ekspektasi dari fungsi *loss*

$$R(K) = \mathrm{E}L(K) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(f_i - f_{K_i} \right)^2 \right)$$

4. Mendapatkan nilai estimasi dari estimator UBR sedemikian sehingga $E\left(\hat{R}(K)\right) = R(K)$ dimana fungsi estimator UBR didefinisikan sebagai berikut:

$$\hat{\mathrm{R}}(\mathrm{K}) = \frac{1}{n}\,||\,\mathbf{(I - A(K))}\underset{\sim}{y}\,||^2 - \frac{\sigma^2}{n}\,\mathrm{tr}\mathbf{(I - A'(K))(I - A(K))} + \frac{\sigma^2}{n}\,\mathrm{tr}\mathbf{(A'(K))(A(K))}$$

b) Mengkaji penurunan metode CV

1. Metode CV membagi data menjadi *n* sub-sampel berukuran *n*-1. Asumsikan $f_{K_{(i)}}$ merepresentasikan fungsi f_{K_i} yang dihitung tanpa observasi ke- *i*. Fungsi $f_{K_{(i)}}$ dibangun dari sub-sampel berukuran *n*-1 yang diambil dari data asli.
2. Jika observasi ke-*i* diperlakukan sebagai observasi tambahan (*holdout sample*), maka fungsi CV adalah sebagai berikut:

$$\mathrm{CV(K)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_{K_{(i)}})^2$$

3. Menguraikan $f_{K_{(i)}}$ sehingga

$$f_{K_{(i)}} = f_{K_i} - s_i(y_i - f_{K_i}) / (1 - s_i)$$

dimana s_i adalah elemen diagonal ke-i dari $\mathbf{A(K)}$.

4. Menaksir fungsi CV sebagai berikut:

$$\text{CV(K)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - f_{K_i}}{1-s_i}\right)^2$$

3.3.2 Studi Simulasi

Sub bab ini menjelaskan langkah-langkah untuk menerapkan spline linier *truncated* pada data simulasi. Penerapan simulasi dilakukan dengan langkah-langkah berikut:

1. Merancang fungsi sebagai berikut:

   $$f(x_{1i,}x_{2i}) = 4 + 2x_{1i} + 3{,}5(x_{1i} - 0{,}5)_+^1 + 1{,}5x_{2i} + 4(x_{2i} - 0{,}6)_+^1$$

2. Merancang model regresi dengan fungsi berikut:

   $$y_i = f(x_{1i,}x_{2i}) + \varepsilon_i, \qquad i = 1,2,\dots,n$$

3. Membangkitkan data sebanyak n untuk masing-masing prediktor komponen nonparametrik $x_{ji}, j = 1,2$ secara *independent* dan berdistribusi Uniform (0,1).
4. Membangkitkan eror random $\varepsilon_i, i = 1,2,\dots,n$ dari distribusi Normal dengan mean 0 dan varians σ^2.
5. Membuat variasi ukuran sampel dengan n yaitu 50 ,75 dan 100 dan varians eror σ^2 yaitu 1, $0{,}75$, $0{,}5$, $0{,}1$ dan $0{,}025$.
6. Melakukan replikasi untuk setiap data bangkitan sebanyak 20 kali.
7. Memodelkan data bangkitan pada langkah (6) dengan menggunakan model spline *truncated* multivariabel.
8. Memilih model dengan metode CV dan UBR.
9. Menghitung nilai MSE dan R^2.
10. Pemodelan dilakukan untuk semua kombinasi n yaitu 50 ,75 dan 100 dan σ^2 yaitu 1, $0{,}75$, $0{,}5$, $0{,}1$ dan $0{,}025$.
11. Membuat analisis berdasarkan *output* yang dihasilkan.

3.3.3 Aplikasi Pada Data Tingkat Pengangguran Terbuka

Tujuan ketiga penelitian ini yaitu membandingkan metode UBR dan CV untuk memilih titik knot optimal pada regresi nonparametrik multivariabel spline *truncated* menggunakan data tingkat pengangguran terbuka di Jawa Tengah. Untuk menjawab tujuan ketiga, dilakukan langkah-langkah sebagai berikut :

1. Membuat *scatter plot* antara masing-masing prediktor dengan respon untuk mengetahui pola hubungan antar variabel.
2. Melakukan uji formal linieritas menggunakan uji Terasvirta.
3. Memodelkan data menggunakan regresi nonparametrik spline *truncated* linier, dimana knot yang digunakan adalah satu knot, dua knot, tiga knot dan kombinasi knot.
4. Memilih titik knot optimal dengan metode UBR.
5. Memilih titik knot optimal dengan metode CV.
6. Mendapatkan model terbaik dengan kriteria nilai MSE dan R^2 minimum.
7. Melakukan uji signifikansi parameter secara simultan.
8. Melakukan uji signifikansi parameter secara parsial.
9. Melakukan pengujian asumsi residual.

(halaman ini sengaja dikosongkan)

BAB 4
HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pembahasan untuk menjawab tujuan 1, 2 dan 3 yang disebutkan dalam Bab 1. Tujuan 1 difokuskan pada kajian teori *Unbiased Risk* (UBR) dan *Cross-Validation* (CV) pada pemilihan titik knot regresi spline *truncated* multivariabel. Tujuan 2 difokuskan pada aplikasi UBR dan CV untuk regresi spline *truncated* pada data simulasi. Tujuan 3 difokuskan pada aplikasi UBR dan CV untuk regresi spline *truncated* pada data Tingkat Pengangguran Terbuka di Jawa Tengah. Hasil pembahasan teori diperoleh berdasarkan tahapan metodologi pada Bab 3, yang hasilnya dijabarkan dalam sub bab-sub bab di bawah ini.

4.1 Kajian Teori *Unbiased Risk* dan *Cross-Validation* pada Regresi Nonparametrik Spline *Truncated* Multivariabel

Pada sub bab ini menyajikan kajian teori untuk menjawab Tujuan 1. Pembahasan pada sub bab 4.1.1 difokuskan pada pembuktian ekspektasi estimator UBR terhadap fungsi *risk*, sedangkan pembahasan pada sub bab 4.1.2 difokuskan pada kriteria CV.

4.1.1 Regresi Nonparametrik Spline *Truncated* Multivariabel

Misalkan diberikan data $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ dimana hubungan antara $(x_{1i}, x_{2i}, \dots, x_{pi})$ dan y_i diasumsikan mengikuti model regresi nonparametrik, maka model regresi nonparametrik adalah sebagai berikut:

$$\begin{aligned} y_i &= f_i + \varepsilon_i \,, i = 1,2, \dots, n \\ &= f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i. \end{aligned} \tag{4.1}$$

Apabila model regresi nonparametrik bersifat aditif, maka:

$$\begin{aligned} y_i &= f(x_{1i}) + f(x_{2i}) + \cdots + f(x_{pi}) + \varepsilon_i \,, i = 1,2, \dots, n \\ &= \sum_{j=1}^{p} f(x_{ji}) + \varepsilon_i. \end{aligned} \tag{4.2}$$

Fungsi $f(x_{ji})$ dihampiri dengan fungsi spline *truncated* multivariabel, sehingga diperoleh model:

$$f(x_{ji}) = \sum_{l=0}^{m} \beta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \beta_{j(m+k)} (x_{ji} - K_{jk})_{+}^{m} \ , i = 1,2,\dots,n; \quad j = 1,2,\dots,p$$

$$= \beta_{j0} + \sum_{l=1}^{m} \beta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \beta_{j(m+k)} (x_{ji} - K_{jk})_{+}^{m}$$

dengan

$$(x_{ji} - K_{jk})_{+}^{m} = \begin{cases} (x_{ji} - K_{jk})^{m} & , x_{ji} \geq K_{jk} \\ 0 & , x_{ji} < K_{jk} \end{cases} .$$

Sehingga, model regresi nonparametrik spline *truncated* multivariabel adalah:

$$y_i = \sum_{j=1}^{p} \beta_{j0} + \sum_{j=1}^{p} \left(\sum_{l=1}^{m} \beta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \beta_{j(m+k)} (x_{ji} - K_{jk})_{+}^{m} \right) + \varepsilon_i \ , i = 1,2,\dots,n$$

dengan mengganti fungsi berikut:

$$\sum_{j=1}^{p} \beta_{j0} = \beta_0$$

maka model regresi nonparametrik spline *truncated* multivariabel menjadi :

$$y_i = \beta_0 + \sum_{j=1}^{p} \left(\sum_{l=1}^{m} \beta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \beta_{j(m+k)} (x_{ji} - K_{jk})_{+}^{m} \right) + \varepsilon_i \ , i = 1,2,\dots,n$$

Atau :

$$\begin{aligned} y_i = {} & \beta_0 + \beta_{11} x_{1i} + \beta_{12} x_{1i}^{2} + \cdots + \beta_{1m} x_{1i}^{m} + \beta_{1(m+1)} (x_{1i} - K_{11})_{+}^{m} \\ & + \beta_{1(m+2)} (x_{1i} - K_{12})_{+}^{m} + \cdots + \beta_{1(m+r)} (x_{1i} - K_{1r})_{+}^{m} + \beta_{21} x_{2i} + \beta_{22} x_{2i}^{2} \\ & + \cdots + \beta_{2m} x_{2i}^{m} + \beta_{2(m+1)} (x_{2i} - K_{21})_{+}^{m} + \beta_{2(m+2)} (x_{2i} - K_{22})_{+}^{m} + \cdots \\ & + \beta_{2(m+r)} (x_{2i} - K_{2r})_{+}^{m} + \cdots + \beta_{p1} x_{pi} + \beta_{p2} x_{pi}^{2} + \cdots + \beta_{pm} x_{pi}^{m} \\ & + \beta_{p(m+1)} (x_{pi} - K_{p1})_{+}^{m} + \beta_{p(m+2)} (x_{pi} - K_{p2})_{+}^{m} + \cdots \\ & + \beta_{p(m+r)} (x_{pi} - K_{pr})_{+}^{m} + \varepsilon_i . \end{aligned}$$

Model regresi spline *truncated* multivariabel dapat disajikan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^m & (x_{11}-K_{11})_+^m & \cdots & (x_{11}-K_{1r})_+^m & \cdots & x_{p1} & \cdots & x_{p1}^m & (x_{p1}-K_{p1})_+^m & \cdots & (x_{p1}-K_{pr})_+^m \\ 1 & x_{12} & \cdots & x_{12}^m & (x_{12}-K_{11})_+^m & \cdots & (x_{12}-K_{1r})_+^m & \cdots & x_{p2} & \cdots & x_{p2}^m & (x_{p2}-K_{p1})_+^m & \cdots & (x_{p2}-K_{pr})_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^m & (x_{1n}-K_{11})_+^m & \cdots & (x_{1n}-K_{1r})_+^m & \cdots & x_{pn} & \cdots & x_{pn}^m & (x_{pn}-K_{p1})_+^m & \cdots & (x_{pn}-K_{pr})_+^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \vdots \\ \beta_{1m} \\ \beta_{1(m+1)} \\ \vdots \\ \beta_{1(m+r)} \\ \vdots \\ \beta_{p1} \\ \vdots \\ \beta_{pm} \\ \beta_{p(m+1)} \\ \vdots \\ \beta_{p(m+r)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dimana :

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X}(\mathbf{K}) = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^m & (x_{11}-K_{11})_+^m & \cdots & (x_{11}-K_{1r})_+^m & \cdots & x_{p1} & \cdots & x_{p1}^m & (x_{p1}-K_{p1})_+^m & \cdots & (x_{p1}-K_{pr})_+^m \\ 1 & x_{12} & \cdots & x_{12}^m & (x_{12}-K_{11})_+^m & \cdots & (x_{12}-K_{1r})_+^m & \cdots & x_{p2} & \cdots & x_{p2}^m & (x_{p2}-K_{p1})_+^m & \cdots & (x_{p2}-K_{pr})_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^m & (x_{1n}-K_{11})_+^m & \cdots & (x_{1n}-K_{1r})_+^m & \cdots & x_{pn} & \cdots & x_{pn}^m & (x_{pn}-K_{p1})_+^m & \cdots & (x_{pn}-K_{pr})_+^m \end{bmatrix}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 & \beta_{11} & \cdots & \beta_{1m} & \beta_{1(m+1)} & \cdots & \beta_{1(m+r)} & \cdots & \beta_{p1} & \cdots & \beta_{pm} & \beta_{p(m+1)} & \cdots & \beta_{p(m+r)} \end{bmatrix}^{\mathrm{T}}, \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$\mathbf{X}(\mathbf{K})$ matriks berukuran $(n \times (1 + p(m + r)))$

$\underset{\sim}{\beta}$ vektor berukuran ($\left(1 + p(m + r)\right) \times 1$)

$\underset{\sim}{y} = \mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ dan $\underset{\sim}{\varepsilon} \sim \text{IIDN}(\underset{\sim}{0}, \sigma^2\mathbf{I})$

Apabila diasumsikan $\underset{\sim}{\varepsilon} \sim \text{N}(\underset{\sim}{0}, \sigma^2\mathbf{I})$, karena $\underset{\sim}{y}$ merupakan kombinasi linier dari $\underset{\sim}{\varepsilon}$, maka $\underset{\sim}{y}$ juga memiliki distribusi normal $\underset{\sim}{y} \sim \text{N}(\text{E}(\underset{\sim}{y}), \text{Var}(\underset{\sim}{y}))$ dengan nilai $\text{E}(\underset{\sim}{y})$ dan $\text{Var}(\underset{\sim}{y})$ sebagai berikut:

$$\begin{aligned} \text{E}(\underset{\sim}{y}) &= \text{E}\,(\mathbf{X}(\mathbf{K})\ \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) \\ &= \mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta} + \text{E}(\underset{\sim}{\varepsilon}) \\ &= \mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta} + \underset{\sim}{0} \\ &= \mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta} \end{aligned} \qquad (4.3)$$

dan

$$\text{Var}(\underset{\sim}{y}) = \text{Var}\,(\mathbf{X}(\mathbf{K})\ \underset{\sim}{\beta} + \underset{\sim}{\varepsilon})$$

$$= 0 + \text{Var}\,(\underset{\sim}{\varepsilon})$$

$$= \sigma^2\mathbf{I}\ . \tag{4.4}$$

Berdasarkan (4.3) dan (4.4), didapatkan $\underset{\sim}{y}$ berdistribusi normal dengan mean $\mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta}$ dan varians $\sigma^2\text{I}$ atau $\underset{\sim}{y} \sim \text{N}(\mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta}, \sigma^2\mathbf{I})$. Salah satu metode untuk mendapatkan nilai estimasi $\underset{\sim}{\beta}$ adalah *Maximum Likelihood Estimation* (MLE). Telah disebutkan bahwa eror berdistribusi normal $\underset{\sim}{\varepsilon} \sim \text{IIDN}(\underset{\sim}{0}, \sigma^2\mathbf{I})$ atau $\varepsilon_i \sim \text{IIDN}(0, \sigma^2)$ sehingga distribusi probabilitas dari ε_i adalah

$$g(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2 \right). \tag{4.5}$$

Berdasarkan persamaan (4.5) diperoleh fungsi likelihood

$$L(\underset{\sim}{\beta}) = \prod_{i=1}^{n} g(\varepsilon_i)$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \varepsilon_i^2 \right)$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\, \underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} \right)$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \left(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}\right)' \left(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}\right) \right). \tag{4.6}$$

Apabila persamaan (4.6) ditransformasi, maka akan didapatkan

$$\ell(\underset{\sim}{\beta}) = \log L(\underset{\sim}{\beta})$$

$$= \log\left(\left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \left(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}\right)' \left(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}\right) \right) \right)$$

$$= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \left(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}\right)' \left(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}\right). \tag{4.7}$$

Untuk mendapatkan estimasi parameter, maka terlebih dahulu dilakukan derivatif parsial. Derivatif parsial terhadap $\underset{\sim}{\beta}$ diberikan oleh:

$$\frac{\partial \ell(\underset{\sim}{\beta})}{\partial \underset{\sim}{\beta}} = \frac{\partial\left(-\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta})'(\underset{\sim}{y} - \mathbf{X(K)}\underset{\sim}{\beta}) \right)}{\partial \underset{\sim}{\beta}}$$

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

$$= -\frac{1}{2\sigma^2}(-2\mathbf{X}'(\mathbf{K})\underset{\sim}{y}+2\mathbf{X}'(\mathbf{K})\mathbf{X}(\mathbf{K})\underset{\sim}{\beta})$$

Apabila derivatif parsial terhadap $\underset{\sim}{\beta}$ sama dengan 0, maka akan didapatkan estimasi parameter $\underset{\sim}{\beta}$, yaitu

$$\underset{\sim}{0} = -\frac{1}{2\sigma^2}(-2\mathbf{X}'(\mathbf{K})\underset{\sim}{y}+2\mathbf{X}'(\mathbf{K})\mathbf{X}(\mathbf{K})\underset{\sim}{\hat{\beta}})$$

$$\underset{\sim}{\hat{\beta}} = (\mathbf{X}'(\mathbf{K})\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}'(\mathbf{K})\underset{\sim}{y} \qquad (4.8)$$

Fungsi penduga dari $f(x)$ adalah

$$\begin{aligned}\hat{f}_K(x) &= \mathbf{X}(\mathbf{K})\,\underset{\sim}{\hat{\beta}}\\ &= \mathbf{X}(\mathbf{K})(\mathbf{X}'(\mathbf{K})\,\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}'(\mathbf{K})\,\underset{\sim}{y}\\ &= \mathbf{A}(\mathbf{K})\,\underset{\sim}{y} \qquad (4.9)\end{aligned}$$

dimana

$$\mathbf{A}(\mathbf{K})=\mathbf{X}(\mathbf{K})(\mathbf{X}'(\mathbf{K})\,\mathbf{X}(\mathbf{K}))^{-1}\mathbf{X}'(\mathbf{K})$$

yang bersifat semi-definit positif dan $\mathbf{X}(\mathbf{K})$ merupakan matriks berukuran $(n \times (1+p(m+r)))$ dari model yang membentuk penduga f_K dan bergantung pada titik knot.

4.1.2 Kajian Metode *Unbiased Risk* (UBR)

Jika diberikan fungsi *loss* sebagai berikut:

$$L(K) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_{K_i} - f_i)^2$$

dimana $\hat{f}_{K_i}$ adalah elemen ke-i dari vektor $\underset{\sim}{f}_K$ berukuran $n \times 1$. Kemudian definisikan fungsi *risk*, yang merupakan nilai ekspektasi dari fungsi *loss*.

$$R(K) = \mathrm{E}(L(K)) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{f}_{K_i} - f_i)^2\right)$$

Berdasarkan persamaan (4.9), $\hat{f}_K = \mathbf{A(K)}\,\underset{\sim}{y}$. Sehingga, fungsi *risk* diuraikan menjadi persamaan berikut:

$$
\begin{aligned}
R(K) &= \frac{1}{n}\mathrm{E}\,\|\,\mathbf{A(K)}(\underset{\sim}{f}+\underset{\sim}{\varepsilon})-\underset{\sim}{f}\,\|^2 \\
&= \frac{1}{n}\mathrm{E}\,\|\,\mathbf{A(K)}\,\underset{\sim}{f}+\mathbf{A(K)}\underset{\sim}{\varepsilon})-\underset{\sim}{f}\,\|^2 \\
&= \frac{1}{n}\mathrm{E}\,\|-(\mathbf{I}-\mathbf{A(K)})\,\underset{\sim}{f}+\mathbf{A(K)}\underset{\sim}{\varepsilon})\,\|^2 \\
&= \frac{1}{n}\|\,(\mathbf{I}-\mathbf{A(K)})\,\underset{\sim}{f}\,\|^2+0+\frac{1}{n}E(\underset{\sim}{\varepsilon}'\mathbf{A}'\mathbf{(K)A(K)}\underset{\sim}{\varepsilon}) \\
&= \frac{1}{n}\|\,(\mathbf{I}-\mathbf{A(K)})\,\underset{\sim}{f}\,\|^2+\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A}'\mathbf{(K))(A(K))} \qquad (4.10)
\end{aligned}
$$

Selanjutnya akan dibuktikan apakah $\mathrm{E}\left(\hat{R}(K)\right) = R(K)$ dimana $\hat{\mathrm{R}}(\mathrm{K})$ merupakan kriteria *unbiased risk* (UBR) dimana nilai $\hat{\mathrm{R}}(\mathrm{K})$ didefinisikan dengan:

$$
\begin{aligned}
\hat{\mathrm{R}}(\mathrm{K}) &= \frac{1}{n}\|\,(\mathbf{I}-\mathbf{A(K)})\underset{\sim}{y}\,\|^2-\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(I-A}'\mathbf{(K))(I-A(K))}+\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A}'\mathbf{(K))(A(K))} \\
&= \frac{1}{n}\underset{\sim}{y}'\mathbf{(I-A}'\mathbf{(K))(I-A(K))}\underset{\sim}{y}-\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(I-A}'\mathbf{(K))(I-A(K))}+\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A}'\mathbf{(K))(A(K))}
\end{aligned}
$$

Kemudian akan dicari nilai ekspektasi dari kriteria UBR.

$$
\begin{aligned}
\mathrm{E}\hat{\mathrm{R}}(\mathrm{K}) &= \mathrm{E}\left(\frac{1}{n}\underset{\sim}{y}'\mathbf{(I-A}'\mathbf{(K))(I-A(K))}\underset{\sim}{y}-\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(I-A}'\mathbf{(K))(I-A(K))}+\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A}'\mathbf{(K))(A(K))}\right) \\
&= \frac{1}{n}\,\mathrm{E}\left(\underset{\sim}{y}'\mathbf{(I-A}'\mathbf{(K))(I-A(K))}\underset{\sim}{y}\right)-\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(I-A}'\mathbf{(K))(I-A(K))} \\
&\quad +\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A}'\mathbf{(K))(A(K))}
\end{aligned}
$$

Diberikan teorema

$$\mathrm{E}(\underset{\sim}{y}'\mathbf{A}\underset{\sim}{y}) = \underset{\sim}{f}'\mathbf{A}\underset{\sim}{f}+\mathrm{tr}\mathbf{(A\Sigma)} \qquad (4.11)$$

karena $\underset{\sim}{y}$ ~N($\mathbf{X}(\mathbf{K})\,\underset{\sim}{\beta}$, $\sigma^2\mathbf{I}$), dengan menggunakan teorema (4.11) didapatkan:

$$
\begin{aligned}
\mathrm{E}(\underset{\sim}{y}'\mathbf{(I-A}'\mathbf{(K))(I-A(K))}\underset{\sim}{y}) &= (\mathbf{X(K)}\underset{\sim}{\beta})'\mathbf{(I-A}'\mathbf{(K))(I-A(K))}(\mathbf{X(K)}\underset{\sim}{\beta}) \\
&\quad +\mathrm{tr}(\mathbf{(I-A}'\mathbf{(K))(I-A(K))}\,\sigma^2\mathbf{I})
\end{aligned}
$$

karena

$$\underset{\sim}{f} = (\mathbf{X(K)}\underset{\sim}{\beta}),$$

maka

$$\mathrm{E}(\underset{\sim}{y}'\mathbf{(I - A'(K))(I - A(K))}\underset{\sim}{y}) = \underset{\sim}{f}'\mathbf{(I - A'(K))(I - A(K))}\underset{\sim}{f} + \sigma^2 \mathrm{tr}\left(\mathbf{(I - A'(K))(I - A(K))}\right)$$

Sehingga nilai ekspektasi dari *unbiased risk* (UBR) adalah:

$$\begin{aligned}
\mathrm{E}\,\hat{\mathrm{R}}(\mathrm{K}) &= \frac{1}{n}\left(\underset{\sim}{f}'\mathbf{(I - A'(K))(I - A(K))}\underset{\sim}{f} + \sigma^2 \mathrm{tr}\left(\mathbf{(I - A'(K))(I - A(K))}\right)\right) \\
&\quad -\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(I - A'(K))(I - A(K))} + \frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A'(K))(A(K))} \\
&= \frac{1}{n}\underset{\sim}{f}'\mathbf{(I - A'(K))(I - A(K))}\underset{\sim}{f} + \frac{\sigma^2}{n}tr\mathbf{(I - A'(K))(I - A(K))} \\
&\quad -\frac{\sigma^2}{n}\mathrm{tr}\mathbf{(I - A'(K))(I - A(K))} + \frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A'(K))(A(K))} \\
&= \frac{1}{n}\underset{\sim}{f}'\mathbf{(I - A'(K))(I - A(K))}\underset{\sim}{f} + \frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A'(K))(A(K))} \\
&= \frac{1}{n}\|\,\mathbf{(I - A(K))}\underset{\sim}{f}\,\|^2 + \frac{\sigma^2}{n}\mathrm{tr}\mathbf{(A'(K))(A(K))} \qquad (4.12)
\end{aligned}$$

Berdasarkan (4.10) dan (4.12), didapatkan bahwa nilai ekspektasi UBR sama dengan fungsi *risk*, sehingga terbukti

$$E\left(\hat{R}(K)\right) = R(K).$$

4.1.3 Kajian Metode *Cross-Validation* (CV)

Metode CV membagi data menjadi n sub-sampel berukuran $n-1$. Asumsikan $f_{K_{(i)}}$ merepresentasikan fungsi f_{K_i} yang dihitung tanpa observasi ke-i. Fungsi $f_{K_{(i)}}$ dibangun dari sub-sampel berukuran $n-1$ yang diambil dari data asli. Jika observasi ke-i diperlakukan sebagai observasi tambahan (*holdout sample*), maka fungsi CV adalah:

$$\mathrm{CV(K)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_{K_{(i)}})^2.$$

Fungsi $f_{K(i)}$ diuraikan sedemikian sehingga

$$f_{K_{(i)}} = f_{K_i} - \frac{s_i(y_i - f_{K_i})}{(1-s_i)} ,$$

dimana s_i adalah elemen diagonal ke-i dari **A**(**K**)**.** Sehingga,

$$\begin{aligned}
\mathrm{CV(K)} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - f_{K_{(i)}})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \left(f_{K_i} - \frac{(y_i - f_{K_i})s_i}{(1-s_i)}\right)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f_{K_i} + \frac{(y_i - f_{K_i})s_i}{(1-s_i)}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(y_i - f_{K_i})(1-s_i)}{(1-s_i)} + \frac{(y_i - f_{K_i})s_i}{(1-s_i)}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(y_i - f_{K_i})(1-s_i+s_i)}{(1-s_i)}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - f_{K_i}}{1-s_i}\right)^2
\end{aligned}$$

sehingga didapatkan kriteria CV yaitu

$$\mathrm{CV(K)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - f_{K_i}}{1-s_i}\right)^2 .$$

4.2 Studi Simulasi

Sub bab berikut menguraikan hasil studi simulasi dari fungsi regresi spline linier *truncated* untuk melihat kesesuaian estimasi titik knot yang dihasilkan oleh metode CV dan UBR. Fungsi yang dirancang untuk studi simulasi ini adalah sebagai berikut:

$y_i = 4 + 2x_{1i} + 3{,}5(x_{1i} - 0{,}5)_+^1 + 1{,}5x_{2i} + 4(x_{2i} - 0{,}6)_+^1 + \varepsilon_i$

Prediktor x_{ji} dibangkitkan secara independen dari distribusi uniform $x_{ji} \sim U(0,1)$, sedangkan eror random $\varepsilon_i, i = 1,2,\ldots,n$ dibangkitkan dari distribusi Normal (0, σ^2). Simulasi dilakukan pada kombinasi varians eror σ^2 yaitu

1, 0,75, 0,5, 0,1 dan 0,025 dan ukuran sampel n yaitu 50 ,75 dan 100 dengan replikasi sebanyak 20 kali. Hasil simulasi berupa rata-rata titik knot untuk metode CV dan UBR disajikan pada Tabel 4.1.

Tabel 4.1 Rata-rata Estimasi Titik Knot dengan CV dan UBR

		n =50		n =75		n =100	
σ^2	knot asli	knot cv	knot ubr	knot cv	knot ubr	knot cv	knot ubr
0,025	0,5	0,49506	0,49663	0,50035	0,50035	0,50794	0,50785
	0,6	0,60757	0,60427	0,59993	0,60158	0,59397	0,59562
0,1	0,5	0,51926	0,52429	0,49501	0,49500	0,48272	0,48272
	0,6	0,57136	0,57298	0,59612	0,58938	0,59282	0,59114
0,5	0,5	**0,44015**	**0,40111**	**0,43073**	**0,46478**	0,49615	0,49463
	0,6	0,59071	**0,56104**	0,57879	0,57877	0,63305	0,63461
0,75	0,5	0,47256	**0,57169**	0,50345	0,47684	0,47738	**0,53634**
	0,6	**0,52095**	0,57515	**0,48297**	**0,50324**	0,57730	**0,56722**
1	0,5	0,53596	**0,55972**	**0,54458**	**0,46329**	0,50520	**0,53234**
	0,6	0,59533	0,56714	0,61222	**0,68521**	0,62720	0,62033

Tabel 4.1 menunjukkan rata-rata estimasi titik knot dari sebanyak 20 replikasi untuk melihat kesesuaian titik knot yang dihasilkan dengan titik knot pada fungsi asli. Kolom yang berwarna biru menunjukkan estimasi titik knot yang lebih mendekati dengan titik knot aslinya, yaitu $K_1 = 0{,}5$ dan $K_2 = 0{,}6$. Pada $n = 50$ untuk semua varians, CV menghasilkan 6 titik knot yang lebih mendekati titik knot asli, sedangkan UBR menghasilkan 6 titik knot yang lebih mendekati titik knot asli. Pada $n = 75$ untuk semua varians, CV menghasilkan 6 titik knot dan UBR 3 titik knot yang lebih mendekati titik knot asli, sedangkan pada $\sigma^2 = 0{,}025$ pada variabel 1, UBR dan CV menghasilkan estimasi knot yang sama, yaitu 0,50035. Pada $n = 100$ untuk semua varians, ada 6 titik knot yang dihasilkan CV dan 3 titik knot yang dihasilkan UBR yang lebih mendekati titik knot asli, juga 1 knot yang sama yang dihasilkan UBR dan CV, yaitu pada $\sigma^2 = 0{,}1$ pada variabel 1. Sebagian besar rata-rata estimasi titik knot yang dihasilkan metode UBR maupun CV sudah mendekati titik knot aslinya. Namun, ada beberapa rata-rata estimasi titik knot yang cukup jauh dari titik knot asli sebagaimana bisa dilihat pada angka yang dicetak tebal. Secara keseluruhan, metode UBR menghasilkan 11 titik knot yang sedikit jauh dari titik knot aslinya,

yaitu dengan selisih di atas 0,035, sedangkah metode CV hanya menghasilkan 5 titik knot yang agak jauh dari titik knot asli.

Dari analisa yang didapat, dapat ditarik kesimpulan bahwa meskipun baik UBR dan CV dapat mengestimasi titik knot yang hampir sama dengan titik knot asli, metode CV lebih banyak menghasilkan titik knot yang mendekati titik knot asli dibandingkan UBR dan estimasi titik knot metode CV yang sedikit jauh dari titik knot asli tidak sebanyak UBR. Sehingga, metode CV lebih sesuai untuk menduga titik knot dibandingkan UBR.

4.3 Aplikasi *Unbiased Risk* (UBR) dan *Cross-Validation* (CV) pada Data Tingkat Pengangguran Terbuka di Jawa Tengah

Pemilihan titik knot menggunakan metode UBR dan CV akan diaplikasikan pada data Tingkat Pengangguran Terbuka (TPT) di Provinsi Jawa Tengah pada tahun 2015. Pemodelan TPT menggunakan model regresi nonparametrik spline *truncated* multivariabel dengan fungsi spline linier. Akan dicobakan titik-titik knot sebanyak satu knot, dua knot, tiga knot, dan kombinasi knot.

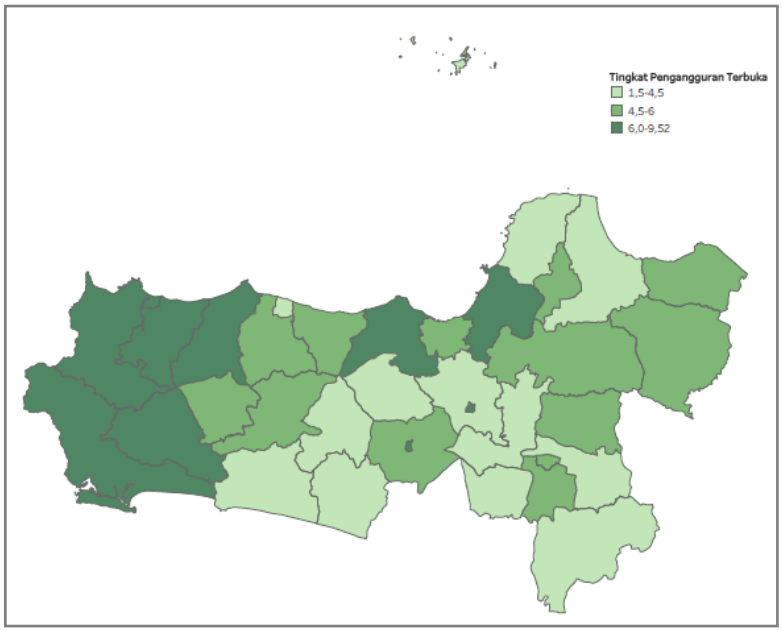
4.3.1 Analisis Deskriptif

Provinsi Jawa Tengah secara administratif dibagi menjadi 29 Kabupaten dan 6 kota. Tingkat pengangguran terbuka di Provinsi Jawa Tengah menduduki peringkat 19 tertinggi dalam skala nasional. Variabel respon yang digunakan adalah Tingkat Pengangguran Terbuka (Y) sedangkan prediktor yang digunakan sebanyak lima variabel yaitu Inflasi (X_1), Laju Pertumbuhan Ekonomi (X_2), Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5). Berikut hasil analisa deskriptif untuk masing-masing variabel penelitian.

Tabel 4.2 Statistik Deskriptif Variabel Respon dan Prediktor

Variabel	Mean	Standar Deviasi	Maksimum	Minimum	Range
Y	4,97	1,73	9,52	1,50	8,02
X_1	2,99	0,58	4,57	1,62	2,95
X_2	5,45	0,47	6,28	3,90	2,38
X_3	91,39	5,83	104,32	77,22	27,10
X_4	1,23	0,13	1,69	1,10	0,59
X_5	68,17	4,02	75,47	60,17	15,30

Berdasarkan tabel 4.2 dapat dilihat karakteristik masing- masing variabel, baik variabel respon maupun prediktor. Variabel respon (Y) yaitu Tingkat Pengangguran Terbuka (TPT) di Jawa Tengah pada Tahun 2015 memiliki nilai rata-rata TPT sebesar 4,97% dengan standar deviasi sebesar 1,73%. Angka TPT tertinggi sebesar 9,52% yang merupakan nilai TPT di Kabupaten Tegal, sedangkan nilai TPT terendah adalah 1,5% yang merupakan nilai TPT di Kabupaten Temanggung. Range nilai TPT di Jawa Tengah dengan 35 kabupaten/kota adalah sebesar 8,02%. Gambar 4.1 merupakan visualisasi kabupaten/kota di Jawa Tengah yang dibagi menjadi tiga kategori TPT: rendah (1,5%-4,5%), sedang (4,5%-6%) dan tinggi (6%-9,52%).



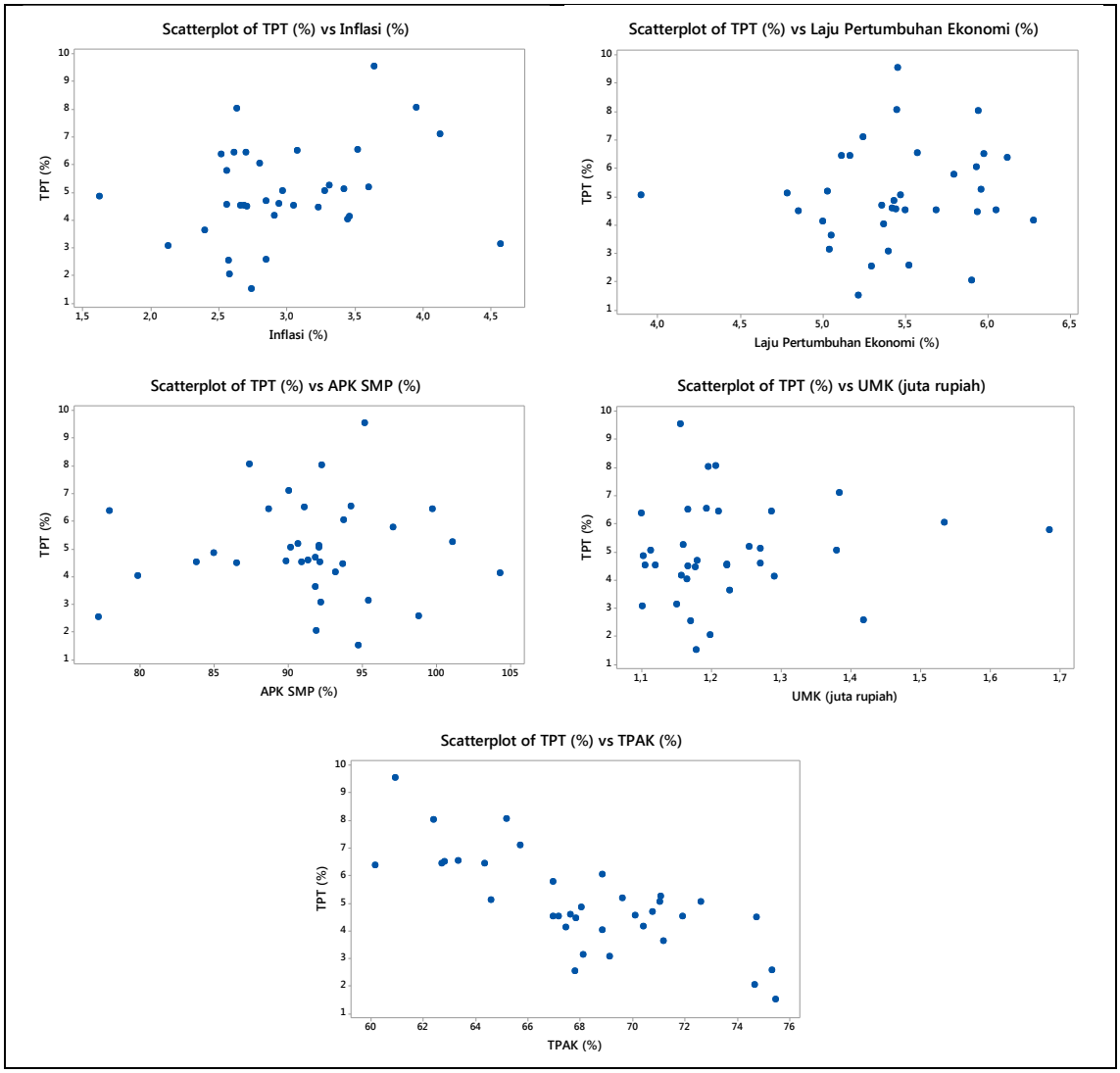
Gambar 4.1 Kabupaten/Kota di Jawa Tengah Berdasarkan Tingkat Pengangguran Terbuka

Dari Gambar 4.1, terlihat bahwa daerah Jawa Tengah bagian barat cenderung memiliki nilai TPT yang cukup tinggi yaitu di atas 6%, sedangkan daerah Jawa Tengah bagian selatan cenderung memiliki TPT yang rendah, yaitu kurang dari 4,5%. Berikut merupakan gambaran karakteristik untuk masing-masing prediktor yang diduga mempengaruhi TPT:

a. Rata-rata angka inflasi di Jawa Tengah pada Tahun 2015 adalah 2,99% dengan standar deviasi sebesar 0,58%. Tingkat inflasi tertinggi di Jawa Tengah terjadi di Kabupaten Jepara, yaitu sebesar 4,57%, sedangkan tingkat inflasi terendah di Kabupaten Purbalingga, yaitu sebesar 1,62% dengan range sebesar 2,95%.
b. Rata-rata tingkat laju pertumbuhan ekonomi di Jawa Tengah pada Tahun 2015 adalah 5,45% dengan standar deviasi sebesar 0,47%. Laju pertumbuhan ekonomi tertinggi terjadi di Kabupaten Kebumen, yaitu sebesar 6,28% dan laju pertumbuhan ekonomi terendah terjadi di Kabupaten Kudus, yaitu 3,9%. Range tingkat laju pertumbuhan ekonomi di Jawa Tengah sebesar 2,38%.
c. Angka Partisipasi Kasar (APK) SMP di Jawa Tengah pada Tahun 2015 memiliki rata-rata 91,39% dengan standar deviasi sebesar 5,83%. APK SMP tertinggi terjadi di Kota Pekalongan, yaitu sebesar 104,32%. Angka APK SMP yang lebih dari 100% berarti adanya penduduk yang belum cukup umur atau melebihi umur 13-15 tahun yang mengikuti pendidikan tingkat SMP. Angka APK SMP terendah terjadi di Kabupaten Klaten dengan angka 77,22% dengan range sebesar 27,1%.
d. Rata-rata Upah Minimum Kabupaten/Kota (UMK) di Jawa Tengah pada Tahun 2015 adalah sebesar 1,23 juta rupiah dengan standar deviasi sebesar 0,13 juta rupiah. Daerah yang memiliki UMK tertinggi adalah Kota Semarang dengan UMK sebesar 1,69 juta rupiah, sedangkan UMK terendah terjadi di Kabupaten Banyumas dengan UMK sebesar 1,1 juta rupiah. Range UMK di Provinsi Jawa Tengah sebesar 0,59 juta rupiah.
e. Rata-rata Tingkat Partisipasi Angkatan Kerja (TPAK) di Provinsi Jawa Tengah pada Tahun 2015 adalah 68,17% dengan standar deviasi sebesar 4,02%. TPAK tertinggi terjadi di Kabupaten Temanggung sebesar 75,47%

dan yang terendah di Kabupaten Banyumas dengan angka 60,17% dengan range sebesar 15,3%.

Tahap awal yang perlu dilakukan sebelum memulai proses pemodelan adalah menyelidiki hubungan antara variabel respon dengan masing-masing prediktor. Pola hubungan antara TPT Jawa Tengah dengan variabel-variabel yang diduga mempengaruhinya dapat dilihat dari *scatter plot* berikut ini:



Gambar 4.2 *Scatter Plot* antara TPT Jawa Tengah Tahun 2015 dengan Variabel-Variabel yang Diduga Mempengaruhi

Berdasarkan Gambar 4.2, dapat dilihat bahwa pola hubungan antara variabel TPT Jawa Tengah Tahun 2015 dengan kelima prediktor, yaitu Inflasi (X_1), Laju

Pertumbuhan Ekonomi (X_2), APK SMP (X_3), UMK (X_4) dan TPAK (X_5) cenderung tidak mengikuti suatu pola tertentu. Untuk mendapatkan keputusan identifikasi yang lebih lanjut apakah suatu model mengikuti pola linier atau nonlinier, dilakukan pengujian linieritas dengan menggunakan Uji Terasvirta. Hipotesis yang digunakan adalah :

H_0 : Antara dan respon cukup dengan model linear

H_1 : Antara prediktor dan respon tidak cukup dengan model linear

Kriteria pengambilan keputusan yaitu tolak H_0 jika *p-value*<0,05. Hasil uji linearitas dengan Uji Terasvirta disajikan dalam Tabel 4.3.

Tabel 4.3 Tabel Hasil Uji Linieritas dengan Uji Terasvirta

Prediktor	*p-value*	Keputusan
X_1	0,1455	Linier
X_2	0,7193	Linier
X_3	0,6113	Linier
X_4	0,9697	Linier
X_5	0,4915	Linier
$X_1 + X_2$	0,0462	Nonlinier
$X_1 + X_2 + X_3$	0,0544	Linier
$X_1 + X_2 + X_3 + X_4$	0,0000	Nonlinier
$X_1 + X_2 + X_3 + X_4 + X_5$	0,0000	Nonlinier

Berdasarkan Tabel 4.3, hubungan antara masing-masing prediktor dengan variabel respon mengikuti hubungan linier. Namun, setelah dilakukan uji bersama-sama, hubungan antara variabel $X_1 + X_2$ dan Y tidak linier. Ketika variabel $X_1 + X_2 + X_3$ dan Y diuji, hubungan kembali linier dengan p-value 0,544. Namun, hubungan antara variabel $X_1 + X_2 + X_3 + X_4$ dan Y tidak linier, begitu pula variabel $X_1 + X_2 + X_3 + X_4 + X_5$ dan Y. Karena regresi mempertimbangkan hubungan antara variabel respon dan prediktor secara serentak, maka berdasarkan Tabel 4.3 hubungan antar antara variabel respon dan kelima variabel preditor adalah nonlinier. Oleh karena itu, data TPT Jawa Tengah dan variabel-variabel yang diduga mempengaruhinya dapat dimodelkan menggunakan regresi nonparametrik spline *truncated* multivariabel.

4.3.2 Pemodelan Tingkat Pengangguran Terbuka di Jawa Tengah Menggunakan Regresi Nonparametrik Spline *Truncated* Linier

Langkah pertama yang dilakukan dalam pemodelan menggunakan regresi nonparametrik spline *truncated* linier multivariabel adalah menentukan titik knot optimal dengan mencobakan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi knot. Pemilihan titik knot optimal menggunakan metode UBR dan CV.

4.3.2.1 Pemilihan Titik Knot Optimal dengan Satu Titik Knot

Pada sub bab ini akan dibahas mengenai pemilihan titik knot optimal pada data Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah dan lima prediktor yang diduga mempengaruhinya. Berikut merupakan model regresi nonparametrik spline *truncated* linier multivariabel dengan satu titik knot menggunakan lima prediktor:

$$\hat{y}_i = \widehat{\beta}_{01} + \widehat{\beta}_{11}x_{1i} + \widehat{\beta}_{12}(x_{1i} - K_{11})_+^1 + \widehat{\beta}_{21}x_{2i} + \widehat{\beta}_{22}(x_{2i} - K_{21})_+^1 + \widehat{\beta}_{31}x_{3i} + \hat{\beta}_{32}(x_{3i} - K_{31})_+^1 + \hat{\beta}_{41}x_{4i} + \hat{\beta}_{42}(x_{4i} - K_{41})_+^1 + \hat{\beta}_{51}x_{5i} + \hat{\beta}_{52}(x_{5i} - K_{51})_+^1$$

Nilai UBR optimal yang dihasilkan dengan menggunakan regresi nonparametrik spline *truncated* linier dengan satu knot disajikan pada Tabel 4.4.

Tabel 4.4 Nilai UBR untuk Spline Linier Satu Knot

Titik Knot Pada					UBR
X_1	X_2	X_3	X_4	X_5	
2,141	4,319	82,002	1,203	62,870	17,976
2,227	4,389	82,799	1,220	63,320	17,985
2,314	4,459	83,596	1,238	63,770	18,037
2,054	4,249	81,205	1,186	62,420	18,071
2,401	4,529	84,394	1,255	64,220	18,123

Berdasarkan Tabel 4.4 didapatkan nilai UBR optimal untuk satu titik knot, yaitu 17,976 pada titik 2,141 untuk variabel X_1, titik 4,319 untuk variabel X_2, titik 82,002 untuk variabel X_3, titik 1,203 untuk variabel X_4, dan titik 62,870 untuk variabel X_5. Selanjutnya akan dicari titik knot optimal menggunakan metode CV dimana hasilnya disajikan pada Tabel 4.5.

Tabel 4.5 Nilai CV untuk Spline Linier Satu Knot

Titik Knot Pada					CV
X_1	X_2	X_3	X_4	X_5	
2,488	4,599	85,191	1,272	64,670	1,985
3,008	5,020	89,973	1,375	67,370	2,171
2,314	4,459	83,596	1,238	63,770	2,184
2,921	4,950	89,176	1,358	66,920	2,314
2,661	4,739	86,785	1,306	65,570	2,336

Berdasarkan Tabel 4.5 didapatkan nilai CV optimal untuk satu titik knot, yaitu 1,985 pada titik 2,488 untuk variabel X_1, titik 4,599 untuk variabel X_2, titik 85,191 untuk variabel X_3, titik 1,272 untuk variabel X_4, dan titik 64,67 untuk variabel X_5.

4.3.2.2 Pemilihan Titik Knot Optimal dengan Dua Titik Knot

Langkah selanjutnya setelah mendapatkan nilai UBR dan CV minimum dengan satu titik knot adalah melakukan pemilihan titik knot optimal dengan dua titik knot. Model regresi nonparametrik spline *truncated* linier multivariabel dengan dua titik knot yang menggunakan lima prediktor adalah sebagai berikut :

$$\hat{y}_i = \widehat{\beta}_{01} + \widehat{\beta}_{11}x_{1i} + \widehat{\beta}_{12}(x_{1i} - K_{11})_+^1 + \widehat{\beta}_{13}(x_{1i} - K_{12})_+^1 + \widehat{\beta}_{21}x_{2i} + \hat{\beta}_{22}(x_{2i} - K_{21})_+^1 + \hat{\beta}_{23}(x_{2i} - K_{22})_+^1 + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}(x_{3i} - K_{31})_+^1 + \hat{\beta}_{33}(x_{3i} - K_{32})_+^1 + \hat{\beta}_{41}x_{4i} + \hat{\beta}_{42}(x_{4i} - K_{41})_+^1 + \hat{\beta}_{43}(x_{4i} - K_{42})_+^1 + \hat{\beta}_{51}x_{5i} + \hat{\beta}_{52}(x_{5i} - K_{51})_+^1 + \hat{\beta}_{53}(x_{5i} - K_{52})_+^1$$

Nilai UBR optimal yang dihasilkan dengan menggunakan regresi nonparametrik spline *truncated* linier dengan dua knot disajikan pada Tabel 4.6

Tabel 4.6 Nilai UBR untuk Spline Linier Dua Knot

	Titik Knot Pada					UBR
	X_1	X_2	X_3	X_4	X_5	
knot 1	3,529	5,440	94,755	1,479	70,070	13,898
knot 2	3,616	5,510	95,552	1,496	70,520	
knot 1	2,314	4,459	83,596	1,238	63,770	14,072
knot 2	3,095	5,090	90,770	1,393	67,820	
knot 1	2,227	4,389	82,799	1,220	63,320	14,073
knot 2	3,095	5,090	90,770	1,393	67,820	
knot 1	3,529	5,440	94,755	1,479	70,070	14,083
knot 2	3,702	5,581	96,349	1,513	70,970	
knot 1	2,314	4,459	83,596	1,238	63,770	14,113
knot 2	3,008	5,020	89,973	1,375	67,370	

Berdasarkan Tabel 4.6 didapatkan nilai UBR optimal untuk dua titik knot, yaitu 13,898 pada titik 3,529 dan 3,616 untuk variabel X_1, titik 5,440 dan 5,510 untuk variabel X_2, titik 94,755 dan 95,552 untuk variabel X_3, titik 1,479 dan 1,496 untuk variabel X_4, dan titik 70,070 dan 70,520 untuk variabel X_5. Selanjutnya akan dicari titik knot optimal menggunakan metode CV dimana hasilnya disajikan pada Tabel 4.7.

Tabel 4.7 Nilai CV untuk Spline Linier Dua Knot

	Titik Knot Pada					CV
	X_1	X_2	X_3	X_4	X_5	
knot 1	2,574	4,669	85,988	1,289	65,120	1,447
knot 2	3,702	5,581	96,349	1,513	70,970	
knot 1	2,488	4,599	85,191	1,272	64,670	1,494
knot 2	3,702	5,581	96,349	1,513	70,970	
knot 1	2,574	4,669	85,988	1,289	65,120	1,564
knot 2	3,355	5,300	93,161	1,444	69,170	
knot 1	2,574	4,669	85,988	1,289	65,120	1,618
knot 2	3,616	5,510	95,552	1,496	70,520	
knot 1	1,707	3,968	78,017	1,117	60,620	1,672
knot 2	2,921	4,950	89,176	1,358	66,920	

Berdasarkan Tabel 4.7 didapatkan nilai CV optimal untuk dua titik knot, yaitu 1,447 pada titik 2,574 dan 3,702 untuk variabel X_1, titik 4,669 dan 5,581 untuk variabel X_2, titik 85,988 dan 96,349 untuk variabel X_3, titik 1,289 dan 1,513 untuk variabel X_4, dan titik 65,120 dan 70,970 untuk variabel X_5.

4.3.2.3 Pemilihan Titik Knot Optimal dengan Tiga Titik Knot

Langkah selanjutnya setelah mendapatkan nilai UBR dan CV minimum dengan dua titik knot adalah melakukan pemilihan titik knot optimal dengan tiga titik knot. Model regresi nonparametrik spline *truncated* linier multivariabel dengan tiga titik knot yang menggunakan lima prediktor adalah sebagai berikut :

$$\hat{y}_i = \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{12}(x_{1i} - K_{11})_+^1 + \hat{\beta}_{13}(x_{1i} - K_{12})_+^1 + \hat{\beta}_{14}(x_{1i} - K_{13})_+^1 + \hat{\beta}_{21}x_{2i} + \hat{\beta}_{22}(x_{2i} - K_{21})_+^1 + \hat{\beta}_{23}(x_{2i} - K_{22})_+^1 + \hat{\beta}_{24}(x_{2i} - K_{23})_+^1 + \hat{\beta}_{31}x_{3i} + \hat{\beta}_{32}(x_{3i} - K_{31})_+^1 + \hat{\beta}_{33}(x_{3i} - K_{32})_+^1 + \hat{\beta}_{34}(x_{3i} - K_{33})_+^1 + \hat{\beta}_{41}x_{4i} + \hat{\beta}_{42}(x_{4i} - K_{41})_+^1 + \hat{\beta}_{43}(x_{4i} - K_{42})_+^1 + \hat{\beta}_{44}(x_{4i} - K_{43})_+^1 + \hat{\beta}_{51}x_{5i} + \hat{\beta}_{52}(x_{5i} - K_{51})_+^1 + \hat{\beta}_{53}(x_{5i} - K_{52})_+^1 + \hat{\beta}_{54}(x_{5i} - K_{53})_+^1$$

Nilai UBR optimal yang dihasilkan dengan menggunakan regresi nonparametrik spline *truncated* linier dengan tiga knot disajikan pada Tabel 4.8.

Tabel 4.8 Nilai UBR untuk Spline Linier Tiga Knot

	Titik Knot Pada					UBR
	X_1	X_2	X_3	X_4	X_5	
knot 1	2,227	4,389	82,799	1,220	63,320	10,207
knot 2	3,702	5,581	96,349	1,513	70,970	
knot 3	3,876	5,721	97,944	1,547	71,870	
knot 1	2,314	4,459	83,596	1,238	63,770	10,227
knot 2	2,835	4,880	88,379	1,341	66,470	
knot 3	3,008	5,020	89,973	1,375	67,370	
knot 1	2,227	4,389	82,799	1,220	63,320	10,231
knot 2	2,835	4,880	88,379	1,341	66,470	
knot 3	2,921	4,950	89,176	1,358	66,920	
knot 1	2,835	4,880	88,379	1,341	66,470	10,238
knot 2	3,095	5,090	90,770	1,393	67,820	
knot 3	3,355	5,300	93,161	1,444	69,170	
knot 1	2,227	4,389	82,799	1,220	63,320	10,247
knot 2	3,702	5,581	96,349	1,513	70,970	
knot 3	3,789	5,651	97,146	1,530	71,420	

Berdasarkan Tabel 4.8 didapatkan nilai UBR optimal untuk tiga titik knot, yaitu 10,207 pada titik 2,227; 3,702 dan 3,876 untuk variabel X_1, titik 4,389; 5,581 dan 5,721 untuk variabel X_2, titik 82,799; 96,349 dan 97,944 untuk variabel X_3, titik 1,220; 1,513 dan 1,547 untuk variabel X_4, dan titik 63,320; 70,970 dan

71,870 untuk variabel X_5. Selanjutnya dicari titik knot optimal menggunakan metode CV yang hasilnya disajikan pada Tabel 4.9.

Tabel 4.9 Nilai CV untuk Spline Linier Tiga Knot

	Titik Knot Pada					CV
	X_1	X_2	X_3	X_4	X_5	
knot 1	1,707	3,968	78,017	1,117	60,620	1,616
knot 2	2,835	4,880	88,379	1,341	66,470	
knot 3	3,616	5,510	95,552	1,496	70,520	
knot 1	2,401	4,529	84,394	1,255	64,220	1,783
knot 2	2,488	4,599	85,191	1,272	64,670	
knot 3	2,748	4,810	87,582	1,324	66,020	
knot 1	1,707	3,968	78,017	1,117	60,620	1,818
knot 2	2,835	4,880	88,379	1,341	66,470	
knot 3	3,702	5,581	96,349	1,513	70,970	
knot 1	2,401	4,529	84,394	1,255	64,220	1,871
knot 2	2,488	4,599	85,191	1,272	64,670	
knot 3	3,095	5,090	90,770	1,393	67,820	
knot 1	2,401	4,529	84,394	1,255	64,220	1,985
knot 2	2,488	4,599	85,191	1,272	64,670	
knot 3	3,616	5,510	95,552	1,496	70,520	

Berdasarkan Tabel 4.9 didapatkan nilai CV optimal untuk tiga titik knot, yaitu 1,616 pada titik 1,707; 2,835 dan 3,616 untuk variabel X_1, titik 3,968; 4,880 dan 5,510 untuk variabel X_2, titik 78,017; 88,379 dan 95,552 untuk variabel X_3, titik 1,117; 1,341 dan 1,496 untuk variabel X_4, dan titik 60,620; 66,470 dan 70,520 untuk variabel X_5.

4.3.2.4 Pemilihan Titik Knot Optimal dengan Kombinasi Satu, Dua dan Tiga Titik Knot

Langkah yang dilakukan setelah mendapatkan nilai UBR dan CV minimum untuk satu, dua dan tiga titik knot adalah melakukan pemilihan titik knot optimal dengan kombinasi titik knot. Nilai UBR optimal yang dihasilkan dengan menggunakan regresi nonparametrik spline *truncated* linier dengan kombinasi titik knot disajikan pada Tabel 4.10.

Tabel 4.10 Nilai UBR untuk Spline Linier Kombinasi Knot

	Titik Knot Pada					UBR
	X_1	X_2	X_3	X_4	X_5	
knot 1	2,227	4,389	82,799	1,220	63,320	
knot 2	3,702	5,581	96,349	1,513	70,970	10,207
knot 3	3,876	5,721	97,944	1,547	71,870	
knot 1	2,227	4,389	94,755	1,220	63,320	
knot 2	3,702	5,581	95,552	1,513	70,970	10,557
knot 3	3,876	5,721	-	1,547	71,870	
knot 1	2,227	5,440	82,799	1,220	63,320	
knot 2	3,702	5,510	96,349	1,513	70,970	10,963
knot 3	3,876	-	97,944	1,547	71,870	
knot 1	2,227	4,389	82,799	1,479	63,320	
knot 2	3,702	5,581	96,349	1,496	70,970	10,974
knot 3	3,876	5,721	97,944	-	71,870	
knot 1	3,529	4,389	82,799	1,220	63,320	
knot 2	3,616	5,581	96,349	1,513	70,970	11,090
knot 3	-	5,721	97,944	1,547	71,870	

Berdasarkan Tabel 4.10 didapatkan nilai UBR optimal untuk tiga titik knot, yaitu 10,207 dengan kombinasi tiga titik knot untuk semua variabel. Nilai knot optimal yang didapatkan menggunakan kombinasi knot sama dengan nilai knot optimal pada tiga titik knot. Selanjutnya akan dicari titik knot optimal menggunakan metode CV dimana hasilnya disajikan pada Tabel 4.11.

Tabel 4.11 Nilai CV untuk Spline Linier Kombinasi Knot

	Titik Knot Pada					CV
	X_1	X_2	X_3	X_4	X_5	
knot 1	2,574	3,968	85,988	1,272	60,620	
knot 2	3,702	4,880	96,349	-	66,470	0,943
knot 3	-	5,510	-	-	70,520	
knot 1	1,707	3,968	85,988	1,272	60,620	
knot 2	2,835	4,880	96,349	-	66,470	0,944
knot 3	3,616	5,510	-	-	70,520	
knot 1	2,574	3,968	85,988	1,289	60,620	
knot 2	3,702	4,880	96,349	1,513	66,470	0,990
knot 3	-	5,510	-	-	70,520	
knot 1	1,707	4,669	85,988	1,272	60,620	
knot 2	2,835	5,581	96,349	-	66,470	1,057
knot 3	3,616	-	-	-	70,520	
knot 1	1,707	3,968	85,988	1,289	60,620	
knot 2	2,835	4,880	96,349	1,513	66,470	1,189
knot 3	3,616	5,510	-	-	70,520	

Berdasarkan Tabel 4.11 didapatkan nilai CV optimal untuk kombinasi titik knot, yaitu 0,943 pada titik 2,574 dan 3,702 untuk variabel X_1, titik 3,968; 4,880 dan 5,510 untuk variabel X_2, titik 85,988 dan 96,349 untuk variabel X_3, titik 1,272 untuk variabel X_4, dan titik 60,620; 66,470 dan 70,520 untuk variabel X_5.

4.3.3 Pemilihan Model Terbaik

Berdasarkan nilai UBR dan CV minimal untuk setiap titik knot yang telah dihitung pada sub bab sebelumnya, berikutnya dilakukan pemilihan model terbaik dengan membandingkan nilai UBR dan CV terkecil dari masing-masing knot yang ditunjukkan pada Tabel 4.12.

Tabel 4.12 Perbandingan Nilai UBR dan CV Minimum

Banyaknya Knot	UBR	CV
1 Titik Knot	17,976	1,985
2 Titik Knot	13,898	1,447
3 Titik knot	10,207	1,616
Kombinasi Titik Knot	10,207	0,943

Berdasarkan Tabel 4.12, nilai UBR minimum adalah 10,207 yang dihasilkan oleh kombinasi titik knot dan tiga titik knot, sedangkan untuk metode CV, nilai minimum dihasilkan oleh kombinasi titik knot sebesar 0,943. Untuk melihat metode mana yang menghasilkan model yang paling baik, dapat dilihat dari nilai *Mean Square Error (MSE)* dan R^2 sebagai kriteria kebaikan model. Nilai MSE dan R^2 untuk metode pemilihan titik knot optimal UBR dan CV disajikan pada Tabel 4.13.

Tabel 4.13 Perbandingan Nilai MSE dan R^2 Metode UBR dan CV

	Nilai Minimum	MSE	R^2
UBR	10,207	0,686	90,590
CV	0,943	0,395	93,047

Berdasarkan Tabel 4.13, terlihat bahwa metode CV menghasilkan nilai MSE yang lebih kecil dan nilai R^2 yang lebih besar dibandingkan dengan metode UBR. Metode CV menghasilkan model dengan 17 parameter, sedangkan metode UBR menghasilkan model dengan 21 parameter. Hal ini berarti metode CV menghasilkan model yang lebih parsimoni dibandingkan metode UBR. Maka dari itu, berdasarkan nilai MSE, R^2 dan model parsimoni yang dihasilkan, metode CV

menghasilkan model yang lebih baik dibandingkan metode UBR.

Langkah selanjutnya adalah mengestimasi parameter β untuk model spline *truncated* linier dengan kombinasi knot dimana nilai estimasi dapat dilihat pada Tabel 4.17. Model regresi nonparametrik spline *truncated* linier multivariabel yang terbentuk adalah sebagai berikut:

$$\hat{y}_i = -20{,}94 + 1{,}67x_{1i} - 0{,}6\,(x_{1i} - 2{,}57)_+^1 - 1{,}64\,(x_{1i} - 3{,}7)_+^1 - 41{,}64\,x_{2i} + 45{,}66(x_{2i} - 3{,}97)_+^1 - 3{,}01(x_{2i} - 4{,}88)_+^1 - 1{,}91\,(x_{2i} - 5{,}51)_+^1 + 0{,}45\,x_{3i} - 0{,}79\,(x_{3i} - 85{,}99)_+^1 + 0{,}55\,(x_{3i} - 96{,}35)_+^1 - 4{,}58\,x_{4i} + 14{,}12\,(x_{4i} - 1{,}27)_+^1 + 2{,}57\,x_{5i} - 3{,}5(x_{5i} - 60{,}62)_+^1 + 1{,}12\,(x_{5i} - 66{,}47)_+^1 - 0{,}63(x_{5i} - 70{,}52)_+^1$$

Model regresi spline *truncated* linier multivariabel dengan kombinasi titik knot 2, 3, 2, 1 dan 3 ini memiliki nilai R^2 sebesar 93,047. Hal ini memiliki arti bahwa model ini dapat menjelaskan Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah sebesar 93,047%.

Nilai estimasi TPT ($\hat{Y}$) dapat dicari menggunakan model terbaik yang sudah didapatkan. Berikut merupakan merupakan perbandingan nilai estimasi menggunakan UBR dan CV yang dapat dilihat pada Tabel 4.14.

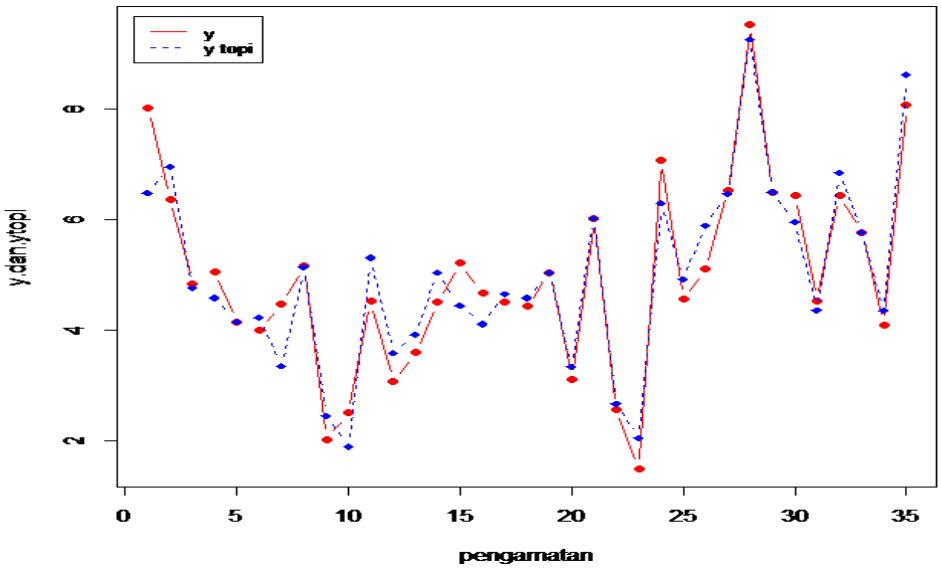
Tabel 4.14 Perbandingan Nilai Estimasi TPT ($\hat{Y}$) Metode UBR dan CV

Pengamatan	Y	$\hat{Y}$	
		Metode UBR	**Metode CV**
Kabupaten Cilacap	8,01	6,473	7,227
Kabupaten Banyumas	6,37	6,945	6,387
Kabupaten Purbalingga	4,84	4,756	4,496
Kabupaten Banjarnegara	5,05	4,582	4,408
Kabupaten Kebumen	4,14	4,155	4,018
Kabupaten Purworejo	4,01	4,225	4,513
Kabupaten Wonosobo	4,47	3,352	4,118
Kabupaten Magelang	5,16	5,137	5,239
Kabupaten Boyolali	2,03	2,456	2,437
Kabupaten Klaten	2,51	1,898	2,090
Kabupaten Sukoharjo	4,52	5,308	4,895
Kabupaten Wonogiri	3,07	3,591	3,840
Kabupaten Karanganyar	3,60	3,919	3,473
Kabupaten Sragen	4,51	5,043	4,794
Kabupaten Grobogan	5,22	4,446	4,470
Kabupaten Blora	4,68	4,112	4,772

Tabel 4.14 Perbandingan Nilai Estimasi TPT ($\hat{Y}$) Metode UBR dan CV (lanjutan)

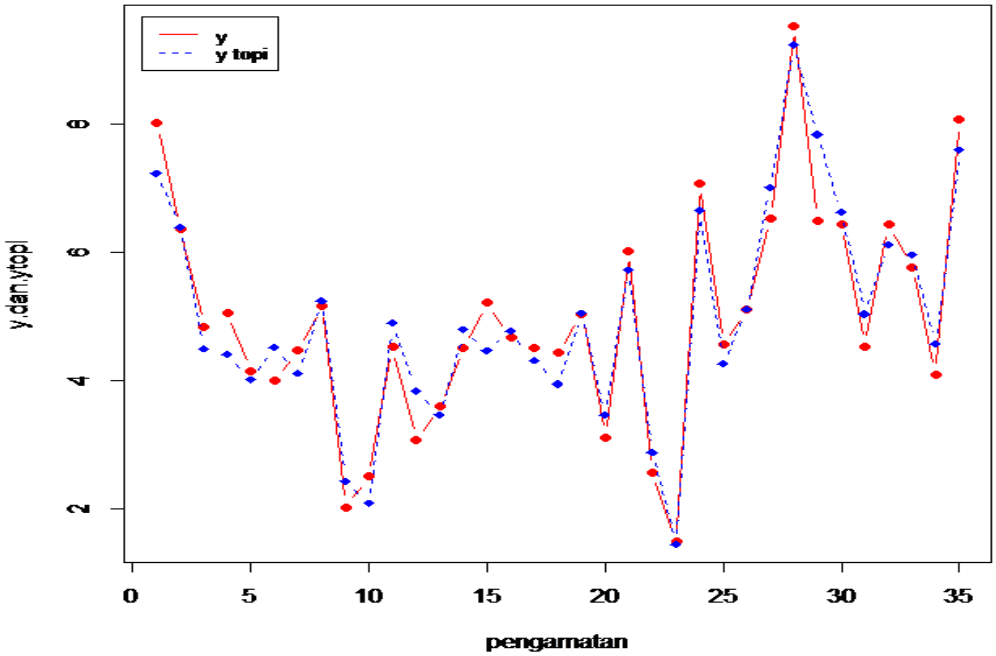
Pengamatan	Y	$\hat{Y}$	
		Metode UBR	**Metode CV**
Kabupaten Rembang	4,51	4,648	4,317
Kabupaten Kudus	5,04	5,040	5,048
Kabupaten Jepara	3,12	3,345	3,461
Kabupaten Demak	6,02	6,020	5,718
Kabupaten Semarang	2,57	2,672	2,880
Kabupaten Temanggung	1,5	2,054	1,447
Kabupaten Kendal	7,07	6,295	6,651
Kabupaten Batang	4,56	4,921	4,259
Kabupaten Pekalongan	5,1	5,887	5,117
Kabupaten Pemalang	6,53	6,464	7,006
Kabupaten Tegal	9,52	9,252	9,227
Kabupaten Brebes	6,49	6,496	7,832
Kota Magelang	6,43	5,951	6,617
Kota Surakarta	4,53	4,359	5,038
Kota Salatiga	6,43	6,849	6,118
Kota Semarang	5,77	5,770	5,953
Kota Pekalongan	4,1	4,354	4,568
Kota Tegal	8,06	8,610	7,590

Perbandingan nilai estimasi TPT ($\hat{Y}$) dan nilai aktual TPT (Y) menggunakan model terbaik yang didapatkan dengan metode UBR dapat digambarkan pada grafik berikut:



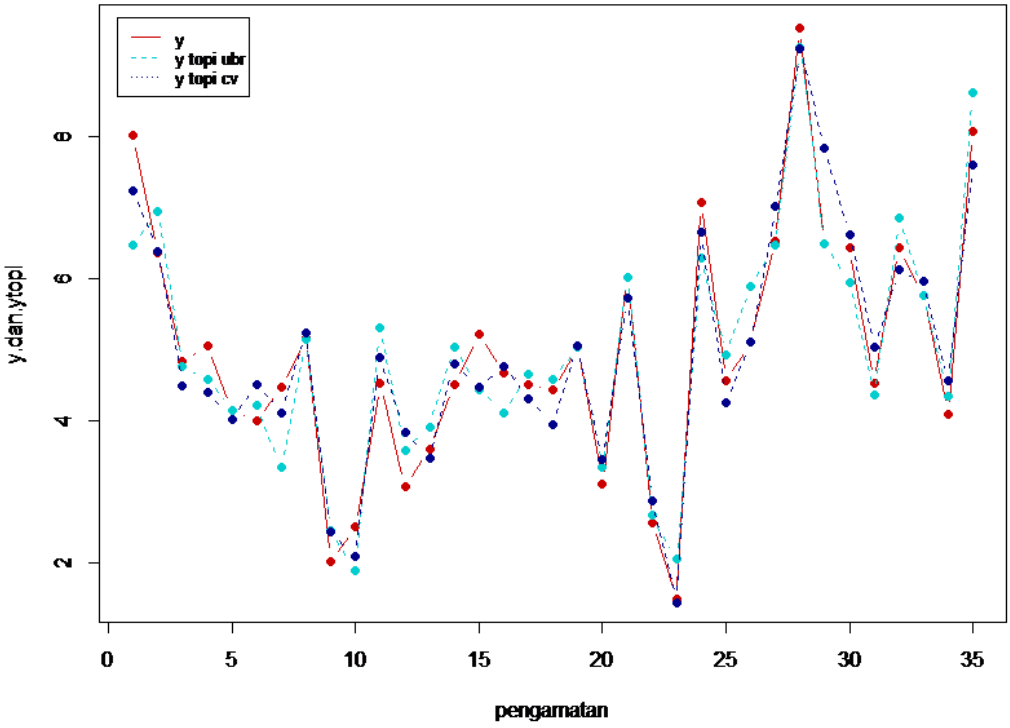
Gambar 4.3 Plot Nilai Estimasi TPT ($\hat{Y}$) dan Nilai Aktual TPT (Y) Menggunakan UBR

Sedangkan perbandingan nilai estimasi TPT ($\hat{Y}$) dan nilai aktual TPT (Y) menggunakan model terbaik yang didapatkan dengan metode CV dapat digambarkan pada grafik berikut:



Gambar 4.4 Plot Nilai Estimasi TPT ($\hat{Y}$) Dan Nilai Aktual TPT (Y) Menggunakan CV

Perbandingan antara nilai estimasi TPT ($\hat{Y}$) dan nilai aktual TPT (Y) menggunakan metode CV dan UBR dapat dilihat pada Gambar 4.5. Terlihat bahwa nilai estimasi menggunakan metode CV yang diwakili oleh garis biru tua lebih mendekati nilai aktual dibandingkan dengan nilai estimasi menggunakan metode UBR yang diwakili oleh garis biru muda.



Gambar 4.5 Plot Gabungan Nilai Estimasi TPT ($\hat{Y}$) dan Nilai Aktual TPT (Y) Menggunakan CV dan UBR

Berdasarkan Gambar 4.3, Gambar 4.4, dan Gambar 4.5, terlihat bahwa nilai estimasi model spline *truncated* multivariabel menggunakan metode pemilihan titik knot CV memberikan estimasi yang relatif mendekati data yang sebenarnya dibandingkan dengan estimasi model spline *truncated* multivariabel menggunakan UBR. Hal ini menunjukkan bahwa untuk data Tingkat Pengangguran Terbuka di Jawa Tengah, pemilihan titik knot menggunakan metode CV menghasilkan model yang lebih baik dibandingkan menggunakan metode UBR.

4.3.4 Pengujian Signifikansi Parameter Secara Simultan

Selanjutnya akan dilakukan pengujian signifikansi parameter secara simultan pada model terbaik yang didapatkan dengan metode CV dan UBR. Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_{11} = \beta_{12} = \cdots = \beta_{p(1+r)} = 0$

H_1: minimal ada satu $\beta_{jl} \neq 0,\ j = 1,2,\ldots,5;\quad l = 1,2,\ldots,1+r$

Hasil ANOVA untuk pengujian hipotesis secara simultan pada model terbaik dengan metode CV disajikan pada Tabel 4.15.

Tabel 4.15 ANOVA Uji Simultan Metode CV

Sumber Variasi	df	*SS*	*MS*	*F* Hitung	P-*Value*
Regresi	16	95,039	5,94		
Error	18	7,102	0,347	15,054	0,000
Total	34	102,142			

Berdasarkan Tabel 4.15 didapatkan nilai *p-value* (0,000) < nilai $\propto$(0,05), sehingga diambil keputusan tolak H_0. Jadi, dapat disimpulkan bahwa terdapat minimal satu parameter β yang siginifikan terhadap model. Sementara itu, hasil ANOVA untuk pengujian hipotesis secara simultan pada model terbaik dengan metode UBR disajikan pada Tabel 4.16.

Tabel 4.16 ANOVA Uji Simultan Metode UBR

Sumber Variasi	df	*SS*	*MS*	*F* Hitung	P-*Value*
Regresi	20	92,531	4,627		
Error	14	9,611	0,687	6,739	0,0003
Total	34	102,140			

Berdasarkan Tabel 4.16 dihasilkan nilai *p-value* (0,0003) < nilai $\propto$(0,05), sehingga diambil keputusan tolak H_0. Jadi, dapat disimpulkan bahwa terdapat minimal satu parameter β yang siginifikan terhadap model. Sehingga, baik model terbaik menggunakan metode CV maupun UBR menghasilkan minimal satu parameter β yang siginifikan terhadap model.

4.3.5 Pengujian Signifikansi Parameter Secara Parsial

Langkah selanjutnya yaitu dilakukan pengujian signifikansi parameter secara parsial pada model terbaik yang didapatkan dengan metode CV dan UBR. Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \beta_{jl} = 0$

$H_1 : \beta_{jl} \neq 0, \quad j = 1,2,\dots,5; \quad l = 1,2,\dots,1+r$

Hasil pengujian signifikansi parameter secara parsial pada model terbaik dengan metode CV disajikan pada Tabel 4.17.

Tabel 4.17 Hasil Uji Signifikansi Parameter Parsial Metode CV

Variabel	Parameter	Estimator	t_{hitung}	*p-value*
Konstan	β_{01}	-20,94	-1,42	0,165
x_1	β_{11}	1,67	2,06	**0,047**
	β_{12}	-0,60	-0,61	0,546
	β_{13}	-1,64	-1,53	0,136
x_2	β_{21}	-41,64	-1,41	0,168
	β_{22}	45,66	1,43	0,161
	β_{23}	-3,01	-1,06	0,295
	β_{24}	-1,91	-1,56	0,128
x_3	β_{31}	0,45	5,73	**0,000**
	β_{32}	-0,79	-6,57	**0,000**
	β_{33}	0,55	4,18	**0,000**
x_4	β_{41}	-4,58	-1,47	0,150
	β_{42}	14,12	3,20	**0,003**
x_5	β_{51}	2,57	1,17	0,250
	β_{52}	-3,50	-1,55	0,129
	β_{53}	1,12	5,06	**0,000**
	β_{54}	-0,63	-3,21	**0,003**

Berdasarkan Tabel 4.17, dari 17 parameter terdapat 7 parameter yang signifikan. Variabel yang berpengaruh terhadap model adalah Inflasi (X_1), Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5), sedangkan variabel Laju Pertumbuhan Ekonomi (X_2) tidak berpengaruh terhadap model. Sementara itu, hasil pengujian signifikansi parameter secara parsial pada model terbaik dengan metode UBR disajikan pada Tabel 4.18.

Tabel 4.18 Hasil Uji Signifikansi Parameter Parsial Metode UBR

Variabel	Parameter	Estimator	t_{hitung}	*p-value*
Konstan	β_{01}	37,18	1,53	0,135
x_1	β_{11}	-0,14	-0,07	0,948
	β_{12}	1,14	0,51	0,612
	β_{13}	10,72	1,66	0,106
	β_{14}	-17,07	-2,03	**0,050**
x_2	β_{21}	-2,62	-0,87	0,388
	β_{22}	2,89	0,78	0,442
	β_{23}	-6,14	-0,82	0,417
	β_{24}	7,47	0,81	0,425

Tabel 4.18 Hasil Uji Signifikansi Parameter Parsial Metode UBR (lanjutan)

Variabel	Parameter	Estimator	t_{hitung}	*p-value*
x_3	β_{31}	0,65	3,08	**0,004**
	β_{32}	-0,76	-2,74	**0,010**
	β_{33}	1,21	1,75	0,089
	β_{34}	-1,42	-1,65	0,108
x_4	β_{41}	3,57	0,53	0,603
	β_{42}	-9,20	-0,93	0,358
	β_{43}	169,84	2,03	**0,050**
	β_{44}	-184,87	-2,04	**0,049**
x_5	β_{51}	-1,22	-3,82	**0,001**
	β_{52}	0,98	2,60	**0,014**
	β_{53}	1,58	1,52	0,138
	β_{54}	-2,09	-1,67	0,105

Berdasarkan Tabel 4.18, terdapat 7 parameter yang signifikan dari total 21 parameter dimana variabel yang berpengaruh terhadap model adalah Inflasi (X_1), Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5), sedangkan variabel Laju Pertumbuhan Ekonomi (X_2) tidak berpengaruh terhadap model. Dengan membandingkan jumlah parameter yang signifikan antara kedua metode, metode CV memiliki proporsi jumlah parameter yang signifikan yang lebih banyak daripada UBR, yaitu 7 parameter yang signifikan dari 17 parameter. Kemudian dilakukan pemeriksaan asumsi residual untuk model dengan metode CV dan UBR. Dari hasil pemeriksaan asumsi residual, kedua model memenuhi asumsi identik, independen dan berdistribusi normal.

4.3.6 Pemilihan Model Terbaik Menggunakan 4 Prediktor

Pada pembahasan sebelumnya, diketahui bahwa variabel Laju Pertumbuhan Ekonomi (X_2) tidak berpengaruh terhadap model, sehingga akan dicari model terbaik dengan memasukkan hanya 4 prediktor, yaitu Inflasi (X_1), Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5). Hasil yang didapat adalah sebagai berikut. Titik knot optimal menggunakan metode UBR hasilnya disajikan pada Tabel 4.19.

Tabel 4.19 Nilai UBR Spline Linier 1 Knot, 2 Knot, 3 Knot dan Kombinasi Knot

	Titik Knot Pada				UBR
	X_1	X_3	X_4	X_5	
1 knot	2,227	82,799	1,220	63,320	19,360
2 knot	3,529	94,755	1,479	70,070	
	3,616	95,552	1,496	70,520	15,872
3 knot	3,529	94,755	1,479	70,070	13,111
	3,789	97,146	1,530	71,420	
	3,876	97,944	1,547	71,870	
Kombinasi knot	3,529	94,755	1,479	70,070	13,111
	3,789	97,146	1,530	71,420	
	3,876	97,944	1,547	71,870	

Titik knot optimal menggunakan metode CV yang hasilnya disajikan pada Tabel 4.20.

Tabel 4.20 Nilai CV Spline Linier 1 Knot, 2 Knot, 3 Knot dan Kombinasi Knot

	Titik Knot Pada				CV
	X_1	X_3	X_4	X_5	
1 knot	2,488	85,191	1,272	64,670	1,446
2 knot	2,748	87,582	1,324	66,020	
	2,835	88,379	1,341	66,470	1,305
3 knot	1,707	78,017	1,117	60,620	
	2,574	85,988	1,289	65,120	1,264
	3,702	96,349	1,513	70,970	
Kombinasi knot	2,748	78,017	1,272	60,620	
	2,835	85,988	-	65,120	1,036
	-	96,349	-	70,970	

Berdasarkan Tabel 4.19 dan 4.20, nilai UBR minimum adalah 13,111 yang dihasilkan oleh tiga titik knot, sedangkan nilai CV minimum sebesar 1,036 yang dihasilkan oleh kombinasi titik knot. Nilai MSE dan R^2 untuk metode pemilihan titik knot optimal UBR dan CV disajikan pada Tabel 4.21.

Tabel 4.21 Perbandingan Nilai MSE dan R^2 Metode UBR dan CV

	Nilai Minimum	**MSE**	$\boldsymbol{R^2}$
UBR	13,111	0,846	85,095
CV	1,036	0,513	89,453

Berdasarkan Tabel 4.21, terlihat bahwa metode CV menghasilkan nilai MSE yang lebih kecil dan nilai R^2 yang lebih besar dibandingkan dengan metode UBR. Metode CV menghasilkan model dengan 14 parameter, sedangkan metode UBR menghasilkan model dengan 17 parameter. Hal ini berarti metode CV

menghasilkan model yang lebih parsimoni dibandingkan metode UBR. Maka dari itu, berdasarkan nilai MSE, R^2 dan model parsimoni yang dihasilkan, metode CV menghasilkan model yang lebih baik dibandingkan metode UBR.

Selanjutnya akan dilakukan pengujian signifikansi parameter secara simultan pada model terbaik yang didapatkan dengan metode CV. Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_{11} = \beta_{12} = \cdots = \beta_{4(1+r)} = 0$

H_1: minimal ada satu $\beta_{jl} \neq 0,\ j = 1,2,\ldots,4; \quad l = 1,2,\ldots,1+r$

Hasil ANOVA untuk pengujian hipotesis secara simultan disajikan pada Tabel 4.22.

Tabel 4.22 ANOVA Uji Simultan Metode CV

Sumber Variasi	df	*SS*	*MS*	*F* Hitung	P-*Value*
Regresi	13	91,369	7,028	13,701	0,000
Error	21	10,773	0,513		
Total	34	102,142			

Berdasarkan Tabel 4.22 didapatkan nilai *p-value* (0,000) < nilai $\propto$(0,05), sehingga diambil keputusan tolak H_0. Jadi, dapat disimpulkan bahwa terdapat minimal satu parameter β yang siginifikan terhadap model. Langkah selanjutnya yaitu dilakukan pengujian signifikansi parameter secara parsial pada model terbaik. Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \beta_{jl} = 0$

$H_1 : \beta_{jl} \neq 0, \qquad j = 1,2,\ldots,4; \quad l = 1,2,\ldots,1+r$

Hasil pengujian signifikansi parameter secara parsial pada model terbaik dengan metode CV disajikan pada Tabel 4.23.

Tabel 4.23 Hasil Uji Signifikansi Parameter Parsial Metode CV

Variabel	**Parameter**	**Estimator**	$\mathbf{t_{hitung}}$	***p-value***
Konstan	β_{01}	0,030	0,994	0,327
x_1	β_{11}	1,510	1,955	0,059
	β_{12}	3,050	0,615	0,543
	β_{13}	-4,194	-0,875	0,388
x_3	β_{21}	-1,328	-1,461	0,153
	β_{22}	1,927	2,029	**0,050**
	β_{23}	-0,969	-6,544	**0,000**
	β_{24}	0,592	3,992	**0,000**

Tabel 4.23 Hasil Uji Signifikansi Parameter Parsial Metode CV (lanjutan)

Variabel	Parameter	Estimator	t_{hitung}	*p-value*
x_4	β_{31}	-4,714	-1,558	0,128
	β_{32}	13,458	3,107	**0,004**
x_5	β_{41}	1,852	1,570	0,126
	β_{42}	-3,030	-2,367	0,024
	β_{43}	1,107	4,346	**0,000**
	β_{44}	-0,323	-1,776	0,085

Berdasarkan Tabel 4.23, terdapat 5 parameter yang signifikan dari total 14 parameter dimana variabel yang berpengaruh terhadap model adalah Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5), sedangkan variabel Inflasi (X_1) tidak berpengaruh terhadap model. Dari hasil pemeriksaan asumsi residual, model memenuhi asumsi identik, independen dan berdistribusi normal.

4.3.7 Pemilihan Model Terbaik Menggunakan 3 Prediktor

Pada pembahasan sebelumnya, diketahui bahwa variabel Inflasi (X_1) tidak berpengaruh terhadap model, sehingga akan dicari model terbaik dengan memasukkan hanya 3 prediktor, yaitu Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5). Titik knot optimal menggunakan metode UBR hasilnya disajikan pada Tabel 4.24.

Tabel 4.24 Nilai UBR Spline Linier Knot 1, 2, 3 dan Kombinasi

	Titik Knot Pada			UBR
	X_3	X_4	X_5	
1 knot	82,002	1,203	62,870	20,833
2 knot	80,408	1,169	61,970	18,050
	81,205	1,186	62,420	
3 knot	80,408	1,169	61,970	15,501
	81,205	1,186	62,420	
	84,394	1,255	64,220	
Kombinasi knot	80,408	1,169	61,970	15,501
	81,205	1,186	62,420	
	84,394	1,255	64,220	

Titik knot optimal menggunakan metode CV yang hasilnya disajikan pada Tabel 4.25.

Tabel 4.25 Nilai CV Spline Linier Knot 1, 2, 3 dan Kombinasi

	Titik Knot Pada			CV
	X_3	X_4	X_5	
1 knot	85,191	1,272	64,670	1,199
2 knot	86,785	1,306	65,570	
	95,552	1,496	70,520	1,119
3 knot	78,017	1,117	60,620	0,893
	86,785	1,306	65,570	
	93,958	1,461	69,620	
Kombinasi knot	86,785	1,117	65,570	0,751
	95,552	1,306	70,520	
	-	1,461	-	

Berdasarkan Tabel 4.24 dan 4.25, nilai UBR minimum adalah 15,501 yang dihasilkan oleh tiga titik knot, sedangkan untuk metode CV, nilai minimum dihasilkan oleh kombinasi titik knot sebesar 0,751. Nilai MSE dan R^2 untuk metode pemilihan titik knot optimal UBR dan CV disajikan pada Tabel 4.26.

Tabel 4.26 Perbandingan Nilai MSE dan R^2 Metode UBR dan CV

	Nilai Minimum	**MSE**	$\boldsymbol{R^2}$
UBR	15,501	0,996	78,550
CV	0,751	0,561	86,811

Berdasarkan Tabel 4.26, terlihat bahwa metode CV menghasilkan nilai MSE yang lebih kecil dan nilai R^2 yang lebih besar dibandingkan dengan metode UBR. Metode CV menghasilkan model dengan 11 parameter, sedangkan metode UBR menghasilkan model dengan 13 parameter. Hal ini berarti metode CV menghasilkan model yang lebih parsimoni dibandingkan metode UBR. Maka dari itu, berdasarkan nilai MSE, R^2 dan model parsimoni yang dihasilkan, metode CV menghasilkan model yang lebih baik dibandingkan metode UBR.

Selanjutnya akan dilakukan pengujian signifikansi parameter secara simultan pada model terbaik yang didapatkan dengan metode CV. Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\beta_{11} = \beta_{12} = \cdots = \beta_{3(1+r)} = 0$

H_1: minimal ada satu $\beta_{jl} \neq 0,\ j = 1,2,\ldots,3; \quad l = 1,2,\ldots,1+r$

Hasil ANOVA untuk pengujian hipotesis secara simultan disajikan pada Tabel 4.27.

Tabel 4.27 ANOVA Uji Simultan Metode CV

Sumber Variasi	df	*SS*	*MS*	*F* Hitung	P-*Value*
Regresi	10	88,671	8,867	15,798	0,000
Error	24	13,471	0,561		
Total	34	102,142			

Berdasarkan Tabel 4.27 didapatkan nilai *p-value* (0,000) < nilai $\propto$(0,05), sehingga diambil keputusan tolak H_0. Jadi, dapat disimpulkan bahwa terdapat minimal satu parameter β yang siginifikan terhadap model. Kemudian dilakukan pengujian signifikansi parameter secara parsial. Hipotesis yang digunakan adalah sebagai berikut:

$H_0 : \beta_{jl} = 0$

$H_1 : \beta_{jl} \neq 0, \quad j = 1,2,\dots,3; \quad l = 1,2,\dots,1+r$

Hasil pengujian signifikansi parameter secara parsial pada Tabel 4.28.

Tabel 4.28 Hasil Uji Signifikansi Parameter Parsial Metode CV

Variabel	**Parameter**	**Estimator**	$\mathbf{t_{hitung}}$	***p-value***
Konstan	β_{01}	-85,64	-2,253	**0,031**
x_3	β_{11}	0,47	6,278	**0,000**
	β_{12}	-0,83	-6,216	**0,000**
	β_{13}	0,57	3,951	**0,000**
x_4	β_{21}	102,28	2,943	**0,006**
	β_{22}	-111,59	-3,020	**0,005**
	β_{23}	29,68	3,287	**0,002**
	β_{24}	-21,31	-2,106	**0,043**
x_5	β_{31}	-0,94	-6,714	**0,000**
	β_{32}	0,88	3,871	**0,000**
	β_{33}	-0,41	-2,191	**0,035**

Berdasarkan Tabel 4.28, semua parameter signifikan dan variabel Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5), berpengaruh terhadap model. Dari hasil pemeriksaan asumsi residual, model memenuhi asumsi identik, independen dan berdistribusi normal. Sehingga model terbaik yang dihasilkan adalah sebagai berikut:

$$\hat{y}_i = -85{,}64 + 0{,}47\, x_{3i} - 0{,}83\, (x_{3i} - 86{,}78)_+^1 + 0{,}57\, (x_{3i} - 95{,}55)_+^1 + 102{,}28\, x_{4i} - 111{,}59\, (x_{4i} - 1{,}12)_+^1 + 29{,}68\, (x_{4i} - 1{,}31)_+^1 - 21{,}31\, (x_{4i} - 1{,}46)_+^1 - 0{,}94\, x_{5i} + 0{,}88 (x_{5i} - 65{,}57)_+^1 - 0{,}41\, (x_{5i} - 70{,}52)_+^1$$

Model regresi spline *truncated* linier multivariabel dengan kombinasi titik knot 2, 3 dan 2 ini memiliki nilai R^2 sebesar 86,81. Hal ini memiliki arti bahwa model ini dapat menjelaskan Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah sebesar 86,81%.

4.3.8 Interpretasi Model Regresi Nonparametrik Spline *Truncated*

Pada sub bab ini akan dilakukan intepretasi model regresi nonparametrik. Berdasarkan analisis yang telah dilakukan, model terbaik yang terbentuk adalah sebagai berikut:

$$\hat{y}_i = -85{,}64 + 0{,}47\, x_{3i} - 0{,}83\, (x_{3i} - 86{,}78)_+^1 + 0{,}57\, (x_{3i} - 95{,}55)_+^1 + 102{,}28\, x_{4i} - 111{,}59\, (x_{4i} - 1{,}12)_+^1 + 29{,}68\, (x_{4i} - 1{,}31)_+^1 - 21{,}31\, (x_{4i} - 1{,}46)_+^1 - 0{,}94\, x_{5i} + 0{,}88(x_{5i} - 65{,}57)_+^1 - 0{,}41\, (x_{5i} - 70{,}52)_+^1$$

Interpretasi model terhadap variabel-variabel yang berpengaruh secara signifikan adalah sebagai berikut:

1. Pengaruh Angka Partisipasi Kasar (APK) SMP (X_3) terhadap TPT apabila variabel lainnya dianggap konstan adalah:

   $$\hat{y}_i = \begin{cases} 0{,}47x_{3i} & , x_{3i} < 86{,}78 \\ -0{,}36x_{3i} + 72{,}03 & , 86{,}78 \le x_{3i} < 95{,}55 \\ 0{,}21x_{3i} + 17{,}56 & , x_{3i} \ge 95{,}55 \end{cases}$$

   Dari model tersebut dapat diintepretasikan yaitu jika suatu daerah memiliki APK kurang dari 86,78%, maka setiap kenaikan APK sebesar 1% mengakibatkan TPT akan cenderung naik sebesar 0,47%. Daerah yang memiliki APK antara 86,78% hingga 95,55%, jika APK naik sebesar 1% maka TPT akan cenderung turun sebesar 0,36%. Sedangkan untuk daerah dengan APK di atas 95,55%, jika APK naik sebesar 1% maka TPT akan cenderung naik sebesar 0,21%.

2. Pengaruh Upah Minimum Kabupaten/Kota (UMK) (X_4) terhadap TPT apabila variabel lainnya dianggap konstan adalah:

   $$\hat{y}_i = \begin{cases} 102{,}28x_{4i} & , x_{4i} < 1{,}12 \\ -9{,}31x_{4i} + 124{,}98 & , 1{,}12 \le x_{4i} < 1{,}31 \\ 20{,}37x_{4i} + 86{,}1 & , 1{,}31 \le x_{4i} < 1{,}46 \\ -0{,}94x_{4i} + 117{,}21 & , x_{4i} \ge 1{,}46 \end{cases}$$

Dari model tersebut dapat diintepretasikan yaitu jika suatu daerah memiliki UMK kurang dari 1,12 juta rupiah, maka setiap kenaikan UMK sebesar 1% mengakibatkan TPT akan cenderung naik sebesar 102,28%. Daerah yang memiliki UMK antara 1,12 juta rupiah hingga 1,31 juta rupiah, jika UMK naik sebesar 1% maka TPT akan cenderung turun sebesar 9,31%. Daerah yang memiliki UMK antara 1,31 juta rupiah hingga 1,46 juta rupiah, jika UMK naik sebesar 1% maka TPT akan cenderung naik sebesar 20,37%. Sedangkan untuk daerah dengan UMK di atas 1,46 juta rupiah, jika UMK naik sebesar 1% maka TPT akan cenderung turun sebesar 0,94%.

3. Pengaruh Tingkat Partisipasi Angkatan Kerja (TPAK) (X_5) terhadap TPT apabila variabel lainnya dianggap konstan adalah:

$$\hat{y}_i = \begin{cases} -0{,}94x_{3i} & , x_{5i} < 65{,}57 \\ -0{,}06x_{3i} - 57{,}70 & ,65{,}57 \leq x_{5i} < 70{,}52 \\ -0{,}47x_{3i} - 28{,}79 & , x_{5i} \geq 70{,}52 \end{cases}$$

Dari model tersebut dapat diintepretasikan yaitu jika suatu daerah memiliki TPAK kurang dari 65,57%, maka setiap kenaikan TPAK sebesar 1% mengakibatkan TPT akan cenderung turun sebesar 0,94%. Daerah yang memiliki TPAK antara 65,57% hingga 70,52%, jika TPAK naik sebesar 1% maka TPT akan cenderung turun sebesar 0,06 %. Sedangkan untuk daerah dengan TPAK di atas 70,52%, jika TPAK naik sebesar 1% maka TPT akan cenderung turun sebesar 0,47%. Hubungan antara TPT dan TPAK berbanding terbalik, hal ini sejalan dengan penelitian Astuti (2017) yang menyatakan TPAK berpengaruh signifikan negatif terhadap TPT di Provinsi Jawa Timur dan penelitian Feng, dkk (2017) yang menyatakan TPAK di Cina turun ketika TPT naik.

(halaman ini sengaja dikosongkan)

BAB 5
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut :

1. a) Kriteria *Unbiased Risk* (UBR) untuk spline *truncated* multivariabel dengan titik knot K yaitu:

$$\hat{R}(K) = \frac{1}{n} || \mathbf{(I - A(K))} \underset{\sim}{y} ||^2 - \frac{\sigma^2}{n} \text{tr}\mathbf{(I - A'(K))(I - A(K))} + \frac{\sigma^2}{n} \text{tr}\mathbf{(A'(K))(A(K))}$$

   b) $E(\hat{R}(K))=R(K)$

   Sehingga, ekspektasi dari estimator UBR sama dengan fungsi *risk*.

2. a) Fungsi *Cross-Validation* (CV) untuk spline *truncated* multivariabel dengan titik knot K yaitu:

$$CV(K) = \frac{1}{n} \sum_{i=1}^{n} (y_i - f_{K_{(i)}})^2$$

   b) Metode CV dapat juga ditulis dengan:

$$CV(K) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - f_{K_i}}{1 - s_i} \right)^2$$

3. Berdasarkan simulasi pada fungsi spline *truncated* dengan eror berdistribusi normal dengan kombinasi variasi ukuran sampel dan varians eror, metode CV lebih banyak menghasilkan titik knot yang mendekati titik knot asli dibandingkan UBR dan juga estimasi titik knot metode CV yang sedikit jauh dari titik knot asli tidak sebanyak UBR. Dapat disimpulkan bahwa metode CV lebih sesuai untuk menduga titik knot dibandingkan UBR.
4. Untuk pemilihan titik knot optimal pada model regresi spline *truncated* multivariabel menggunakan UBR dan CV pada aplikasi data Tingkat Pengangguran Terbuka (TPT) di Jawa Tengah Tahun 2015, model terbaik untuk metode CV menghasilkan nilai MSE=0,561, R^2=86,811 dan model optimal dengan 11 parameter, sedangkan UBR menghasilkan MSE=0,996, R^2=78,550 dan model optimal dengan 13 parameter. CV menghasilkan nilai

MSE yang lebih kecil, R^2 yang lebih besar, dan model yang parsimoni dibandingkan dengan metode UBR. Sehingga dapat diambil kesimpulan bahwa metode CV menghasilkan model yang lebih baik dibandingkan metode UBR.

Model terbaik dengan metode CV yang menggunakan kombinasi titik knot 2-3-2 adalah sebagai berikut:

$$\hat{y}_i = -85{,}64 + 0{,}47\, x_{3i} - 0{,}83\, (x_{3i} - 86{,}78)_+^1 + 0{,}57\, (x_{3i} - 95{,}55)_+^1 + 102{,}28\, x_{4i} - 111{,}59\, (x_{4i} - 1{,}12)_+^1 + 29{,}68\, (x_{4i} - 1{,}31)_+^1 - 21{,}31\, (x_{4i} - 1{,}46)_+^1 - 0{,}94\, x_{5i} + 0{,}88(x_{5i} - 65{,}57)_+^1 - 0{,}41\, (x_{5i} - 70{,}52)_+^1$$

Dari kelima prediktor yang dianalisis, variabel Angka Partisipasi Kasar SMP (X_3), Upah Minimum Kabupaten/Kota (X_4) dan Tingkat Partisipasi Angkatan Kerja (X_5) berpengaruh terhadap TPT di Provinsi Jawa Tengah.

5.2. Saran

1. Penelitian ini terbatas pada data *cross-section*, pada penelitian selanjutnya dapat membandingkan metode UBR dan CV menggunakan data longitudinal.
2. Simulasi yang dilakukan dalam penelitian ini terbatas pada data dengan residual berdistribusi normal, sehingga pada penelitian selanjutnya dapat membandingkan metode UBR dan CV pada data simulasi dengan residual yang berdistribusi selain normal.
3. Perlu dilakukan kajian metode pemilihan titik knot yang lain pada data TPT seperti GML sehingga dapat dilakukan perbandingan yang lebih komprehensif.

DAFTAR PUSTAKA

Astuti, W.I., Ratnasari, V. dan Wibowo, W., 2017, Analisis Faktor yang Berpengaruh Terhadap Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Menggunakan Regresi Data Panel. *Jurnal Sains dan Seni ITS*, 6 (1), 144-149.

Badan Pusat Statistik, 2007, *Analisis Perkembangan Statistik Ketenagakerjaan (Laporan Sosial Indonesia 2007)*, BPS, Jakarta.

Badan Pusat Statistik, 2015, *Profil Ketenagakerjaan Jawa Tengah Hasil Sakernas Agustus 2015*, BPS, Jawa Tengah.

Badan Pusat Statistik, 2015, *Statistik Pendidikan Jawa Tengah 2015*, BPS, Jawa Tengah.

Badan Pusat Statistik, 2016, "Indeks Harga Konsumen dan Inflasi Jawa Tengah Menurut Kabupaten/Kota, 2008-2016", [online], (https://jateng.bps.go.id/linkTableDinamis/view/id/41, diunduh tanggal 7 September 2017).

Badan Pusat Statistik, 2016, "Laju Pertumbuhan PDRB Atas Dasar Harga Konstan 2010 Menurut Kabupaten/Kota di Jawa Tengah (Persen), 2010 - 2016", [online], (https://jateng.bps.go.id/linkTabelStatis/view/id/1413, diunduh tanggal 7 September 2017).

Badan Pusat Statistik, 2017, *Provinsi Jawa Tengah Dalam Angka 2017*, BPS, Jawa Tengah.

Badan Pusat Statistik, 2017, "Keadaan Ketenagakerjaan Jawa Tengah, Februari 2017: Tingkat Pengangguran Terbuka Sebesar 4,15 Persen", [online], (https://jateng.bps.go.id/website/brs_ind/brsInd-20170505153354.pdf, diunduh tanggal 7 September 2017).

Badan Pusat Statistik, "Angka Partisipasi Kasar (APK)", [online], (https://sirusa.bps.go.id/index.php?r=indikator/view&id=8, diakses tanggal 27 Agustus 2017).

Badan Pusat Statistik, "Konsep Inflasi dan Indeks Harga Konsumen", [online], (https://jateng.bps.go.id/Subjek/view/id/3, diakses tanggal 2 Agustus 2017).

Bank Indonesia, "Definisi Inflasi", [online], (http://www.bi.go.id/id/moneter/inflasi/pengenalan/Contents/Default.aspx, diakses tanggal 31 Juli).

Bank Indonesia, "Disagregasi Inflasi", [online], (http://www.bi.go.id/id/moneter/inflasi/pengenalan/Contents/Disagregasi.aspx, diakses tanggal 31 Juli 2017).

Bank Indonesia, “Pentingnya Kestabilan Harga”, [online], (http://www.bi.go.id/id/moneter/inflasi/pengenalan/Contents/Pentingnya.aspx, diakses tanggal 31 Juli 2017).

Bhattarai, K., 2016, “Unemployment-Inflation Trade-offs in OECD Countries”, *Economic Modelling*, 58, 93-103.

Biro Analisa Anggaran dan Pelaksanaan APBN, “Analisis Keberadaan Trade Off Inflasi dan Pengangguran (Kurva Phillips) di Indonesia”, [online], (http://www.dpr.go.id/doksetjen/dokumen/apbn_analisis_keberadaan_tradeoff_inflasi_dan_pengangguran_%28kurva_phillips%29_di_indonesia20140821142142.pdf, diakses tanggal 10 September 2017)

Budiantara, I. N., 2004, Model Spline Multivariabel dalam Regresi Nonparametrik, *Makalah Seminar Nasional Matematika*, Jurusan Matematika ITS, Surabaya.

Budiantara, I. N., 2005, Penentuan Titik-Titik Knots dalam Regresi Spline , *Jurnal Jurusan Statistika FMIPA-ITS*, Surabaya.

Budiantara, I.N., Lestari, B., Islamiyati, A., dan Wibowo, W., 2009,“Pemilihan Knot Optimal Dalam Estimator Spline Terbobot Pada Regresi Nonparametrik Heteroskedastik Data Longitudinal”, Prosiding Seminar Nasional Statistika XI, Institut Teknologi Sepuluh Nopember, Surabaya.

Budiantara, I.N., Suryadi, F., Otok, B.W., dan Guritno, S., 2006. Pemodelan B-Spline dan MARS Pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK. PETRA Surabaya. *Jurnal Teknik Industri*. 8(1): 1-13.

Craven, P., dan Wahba, G., 1979, "Smoothing Noisy Data With Spline Functions," *Numerische Mathematik*, 31, 377-403.

Eubank, R.L., 1999, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.

Feng, S., Hu, Y. dan Moffit, R., 2017, “Long Run Trends in Unemployment and Labor Force Participation in China”, *Journal of Comparative Economics*, 45(2), 304-324.

Fox, J., 2008, *Applied Regression Analysis and Generalized Linier Models*, Sage Publications, London.

Gorry, A., 2013, “Minimum Wages and Youth Unemployment”, *European Economic Review*, 64, 57-75

Gujarati, D., 2003, *Ekonometrika Dasar*, Terjemahan: Sumarno Zain, Erlangga, Jakarta.

Gu, C.,1992, “Cross-Validating Non-Gaussian Data”, *Journal of Computational and Graphical Statistics*, Vol.1, No.2, 169-179.

Ha¨rdle, W., 1990, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Haug, A. dan King, I., 2014, "In The Long Run, US Unemployment Follows Inflation Like A Faithful Dog", *Journal of Macroeconomics*, 41, 42-52.

Jhingan, 1993, *Ekonomi Pembangunan dan Perencanaan*, Rajawali Pers, Jakarta.

Kartasapoetra, G., 1992, *Kalkulasi Pengendalian Biaya Produksi*, Rineka Cipta, Jakarta.

Kohn, R., Ansley, C. F., dan Tharm, D., 1991, "The Performance of Cross-Validation and Maximum Likelihood Estimators of Spline Smoothing Parameters," *Journal of the American Statistical Association*, 86, 1042- 1050.

Lavrinovicha, I., Lavrinenko, O. dan Treinovskis, J., 2015, "Influence of Education on Unemployment Rate and Incomes of Residents", *Procedia-Social and Behavioral Sciences*, 174, 3824-3831.

Montoya, E., Ulloa, N. dan Miller, V., 2014, "A Simulation Study Comparing Knot Selection Methods With Equally Spaced Knots in a Penalized Regression Spline", *International Journal of Statistics and Probability*,pp. Vol. 3, No. 3.

Muslim, M., 2014, "Pengangguran Terbuka dan Determinannya", *Jurnal Ekonomi dan Studi Pembangunan*, 15, 171-181.

Opsomer, J., Wang, Y. dan Yang, Y., 2001, "Nonparametric Regression with Correlated Errors", *Statistical Science*, Vol. 16, No.2, 134-153.

Osterholm, P., 2010, "Unemployment and Labour-Force Participation in Sweden", *Economics Letters*, Vol. 106, 205-208.

Prahutama, A., 2013,"Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur", Prosiding Seminar Nasional Statistika, Universitas Diponegoro, Semarang.

Rencher,A.C. dan Scaalje, G.B.,2007, *Linear Models in Statistics*, A John Wiley and Son Inc. Publication, New Jersey.

Sari, R.S. dan Budiantara, I.N., 2012, Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Regresi Spline Multivariabel. *Jurnal Sains dan Seni ITS*, 1 (1), 236-241.

Slonimczyk, F. dan Skott, P., 2012, "Employment and Distribution Effects of The Minimum Wage", *Journal of Economic Behavior and Organization*, 84, 245-264.

Snieska, V., Valodkiene, G., Daunoriene, A., dan Draksaite, A., 2015, "Education and Unemployment in European Union Economic Cycles", *Procedia-Social and Behavioral Sciences*, 213, 211-216.

Sujianto, R., 2017, “Jateng Semakin Diminati Investor Untuk Pendirian Industri”, [online], (https://semarangpedia.com/jateng-semakin-diminati-investor-untuk-pendirian-industri/, diakses tanggal 7 September 2017).

Sukirno, S., 2004, *Pengantar Teori Makroekonomi*, PT Raja Grafindo Persada, Jakarta.

Wahba, G., 1990, *Spline Model for Observational Data*, SIAM, XII, Philadelphia.

Wang, Y., 1998, “Spline Smoothing Models with Correlated Error”, *Journal of the Royal Statistical Sociaty, Series B, 5r0, 341-348.*

Wang,Y., 2011, *Smoothing Splines Methods and Applications*, CRC Press, California.

Wu, H. dan Zhang, J.T, 2006, *Nonparametric Regression Methods for Longitudinal Data Analysis*, A John Wiley and Son Inc. Publication, New Jersey.

Lampiran 1 : Data Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah dan Faktor-Faktor yang Diduga Berpengaruh.

No	Pengamatan	Y	X_1	X_2	X_3	X_4	X_5
1	Kabupaten Cilacap	8,01	2,63	5,94	92,31	1,196	62,39
2	Kabupaten Banyumas	6,37	2,52	6,12	77,97	1,100	60,17
3	Kabupaten Purbalingga	4,84	1,62	5,43	85	1,102	68,05
4	Kabupaten Banjarnegara	5,05	2,97	5,47	92,12	1,113	72,61
5	Kabupaten Kebumen	4,14	2,91	6,28	93,24	1,158	70,43
6	Kabupaten Purworejo	4,01	3,45	5,37	79,84	1,165	68,86
7	Kabupaten Wonosobo	4,47	2,71	4,85	86,55	1,166	74,75
8	Kabupaten Magelang	5,16	3,6	5,03	90,67	1,255	69,63
9	Kabupaten Boyolali	2,03	2,58	5,91	91,91	1,198	74,68
10	Kabupaten Klaten	2,51	2,57	5,30	77,22	1,170	67,79
11	Kabupaten Sukoharjo	4,52	2,69	5,69	83,84	1,223	67,17
12	Kabupaten Wonogiri	3,07	2,13	5,40	92,25	1,101	69,12
13	Kabupaten Karanganyar	3,6	2,4	5,05	91,85	1,226	71,18
14	Kabupaten Sragen	4,51	3,05	6,05	90,93	1,105	71,91
15	Kabupaten Grobogan	5,22	3,31	5,96	101,12	1,160	71,09
16	Kabupaten Blora	4,68	2,85	5,36	91,83	1,180	70,77
17	Kabupaten Rembang	4,51	2,66	5,50	92,19	1,120	66,97
18	Kabupaten Pati	4,43	3,23	5,94	93,69	1,177	67,83
19	Kabupaten Kudus	5,04	3,28	3,90	90,17	1,380	71,04
20	Kabupaten Jepara	3,12	4,57	5,04	95,44	1,150	68,13
21	Kabupaten Demak	6,02	2,8	5,93	93,78	1,535	68,84
22	Kabupaten Semarang	2,57	2,85	5,52	98,82	1,419	75,33
23	Kabupaten Temanggung	1,5	2,74	5,21	94,74	1,178	75,47
24	Kabupaten Kendal	7,07	4,13	5,25	90,05	1,383	65,7
25	Kabupaten Batang	4,56	2,94	5,42	91,33	1,270	67,62
26	Kabupaten Pekalongan	5,1	3,42	4,78	92,07	1,271	64,6
27	Kabupaten Pemalang	6,53	3,52	5,58	94,24	1,193	63,32
28	Kabupaten Tegal	9,52	3,64	5,45	95,22	1,155	60,91
29	Kabupaten Brebes	6,49	3,08	5,98	91,1	1,167	62,81
30	Kota Magelang	6,43	2,7	5,11	88,68	1,211	64,35
31	Kota Surakarta	4,53	2,56	5,44	89,88	1,222	70,12
32	Kota Salatiga	6,43	2,61	5,17	99,74	1,287	62,7
33	Kota Semarang	5,77	2,56	5,80	97,12	1,685	66,96
34	Kota Pekalongan	4,1	3,46	5,00	104,32	1,291	67,47
35	Kota Tegal	8,06	3,95	5,45	87,39	1,206	65,17

Keterangan :

Y : Tingkat Pengangguran Terbuka (dalam persen)

X_1 : Inflasi (dalam persen)

X_2 : Laju Pertumbuhan Ekonomi (dalam persen)

X_3 : Angka Partisipasi Kasar SMP (dalam persen)

X_4 : Upah Minimum Kabupaten/Kota (dalam juta rupiah)

X_5 : Tingkat Partisipasi Angkatan Kerja (dalam persen)

Lampiran 2 : Sintaks R Fungsi Utama untuk Memanggil Fungsi Pendukung Lainnya dan Mengeluarkan Output Keseluruhan.

```
spline.truncated.linier.multivariabel.ubr.cv=function(x,y,b,taraf.alpha){
 #fungsi utama spline linier multivariabel
 #fungsi ini memanggil fungsi fungsi yang lain dalam membentuk matriks truncated, penaksiran parameter, dan
mencari gcv minimum
 #b adalah banyaknya titik yang dicoba sebagai knot untuk masing masing variabel dalam trial error pencarian
knot optimum
 source("hasilperhitungan.R")
 source("kombinasiknotubr.R")
 source("kombinasiknotcv.R")

 cv=kombinasi.knot.cv(x,y,b)
 cvsatuknot=cv$cv.1knot
 cv.satuknot=cv$satuknot.optimum
 cvduaknot=cv$cv.2knot
 cv.duaknot=cv$duaknot.optimum
 cvtigaknot=cv$cv.3knot
 cv.tigaknot=cv$tigaknot.optimum
 cv.knot.optimum=cv$knot.optimum
 cv.minimum=cv$cv.minimum
 mse.cv=hitung(x,y,cv.knot.optimum,taraf.alpha)$MSE

 ubr=kombinasi.knot.ubr(x,y,b)
 ubrsatuknot=ubr$ubr.1knot
 ubr.satuknot=ubr$satuknot.optimum
 ubrduaknot=ubr$ubr.2knot
 ubr.duaknot=ubr$duaknot.optimum
 ubrtigaknot=ubr$ubr.3knot
 ubr.tigaknot=ubr$tigaknot.optimum
 ubr.knot.optimum=ubr$knot.optimum
 ubr.minimum=ubr$ubr.minimum
 mse.ubr=hitung(x,y,ubr.knot.optimum,taraf.alpha)$MSE

 gabungan=c(mse.ubr,mse.cv)
 indeks=which.min(gabungan)
 if(indeks==1){
  kriteria="UBR"
  knot.optimum=ubr.knot.optimum
  kriteria.minimum=ubr.minimum
  kriteria.1knot=ubrsatuknot
  satuknot=ubr.satuknot
  kriteria.2knot=ubrduaknot
  duaknot=ubr.duaknot
  kriteria.3knot=ubrtigaknot
  tigaknot=ubr.tigaknot
 } else {
  kriteria="CV"
  knot.optimum=cv.knot.optimum
  kriteria.minimum=cv.minimum
  kriteria.1knot=cvsatuknot
  satuknot=cv.satuknot
```

Lampiran 2 : Sintaks R Fungsi Utama untuk Memanggil Fungsi Pendukung Lainnya dan Mengeluarkan Output Keseluruhan (lanjutan)

```
  kriteria.2knot=cvduaknot
  duaknot=cv.duaknot
  kriteria.3knot=cvtigaknot
  tigaknot=cv.tigaknot
 }

 beta=hitung(x,y,knot.optimum,taraf.alpha)$beta
 tabel.anova=hitung(x,y,knot.optimum,taraf.alpha)$tabel.anova
 rsquare=hitung(x,y,knot.optimum,taraf.alpha)$R.square
 mse=hitung(x,y,knot.optimum,taraf.alpha)$MSE
 inferensi=hitung(x,y,knot.optimum,taraf.alpha)$inferensi
 ytopi=hitung(x,y,knot.optimum,taraf.alpha)$ytopi
 residual=y-ytopi
 plot(y,xlab="pengamatan",ylab="y.dan.ytopi",type="b",pch=19,col="red")
 lines(ytopi,col="blue",type="b",pch=18,lty=3)
 legend("topleft", legend=c("y", "y topi"),inset=.02,
       col=c("red", "blue"), lty=1:2, cex=0.8)

 kenormalan=ks.test(residual,"pnorm",mean=mean(residual),sd=sd(residual))
 if(kenormalan$p.value >= taraf.alpha){
  keputusan.kenormalan=c("residual berdistribusi normal")
  }
 else { keputusan.kenormalan=c("residual tidak berdistribusi normal")
 }
 gleyjser=hitung(x,abs(residual),knot.optimum,taraf.alpha)$tabel.anova

uji.kenormalan.KS=data.frame(D=as.numeric(kenormalan$statistic),p.value=kenormalan$p.value,keputusan=keput
usan.kenormalan)
 durbin=durbinWatsonTest(as.vector(residual))
 matrikstruncated=mtruncated(x,knot.optimum)$xtruncated
 p.value.dw=durbinWatsonTest(lm(y~matrikstruncated[,-1]))$p
 if (as.numeric(paste(gleyjser[1,6]))>=taraf.alpha){
  kesimpulan.gly=c("Tidak terjadi heterokedastisitas")
 }
 else {
  kesimpulan.gly=c("Terjadi heterokedastisitas")
 }
 if(p.value.dw >= taraf.alpha){
  keputusan.dw=c("gagal tolak Hnol")
  kesimpulan.dw=c("Tidak terjadi autokorelasi")
  }
 else { keputusan.dw=c("tolak Hnol")
 kesimpulan.dw=c("Ada autokorelasi")
 }
 cat("==================================================","\n")
 cat("PEMILIHAN TITIK KNOT OPTIMAL MENGGUNAKAN CV DAN UBR","\n")
 cat("==================================================","\n")
 cat("Nilai MSE CV : ",mse.cv,"\n")
 cat("Nilai MSE UBR : ",mse.ubr,"\n")
```

Lampiran 2 : Sintaks R Fungsi Utama untuk Memanggil Fungsi Pendukung Lainnya dan Mengeluarkan Output Keseluruhan (lanjutan)

```
cat("=====================================================","\n")
 cat("Kriteria minimum dihasilkan oleh",kriteria,"\n")
 cat("","\n")
 cat("","\n")
 cat("=====================================================","\n")
 cat("satu knot optimum untuk masing-masing variabel","\n")
 cat("=====================================================","\n")
 cat("","\n")
 cat(kriteria,"minimum = ",kriteria.1knot,"\n")
 cat("","\n")
 for (i in 1:length(satuknot)){
  cat("knot optimum variabel ke-",i,"\n")
  cat(satuknot[[i]],"\n")
  cat("","\n")
 }
 cat("=====================================================","\n")
 cat("dua knot optimum untuk masing-masing variabel","\n")
 cat("=====================================================","\n")
 cat("","\n")
 cat(kriteria,"minimum = ",kriteria.2knot,"\n")
 cat("","\n")
 for (i in 1:length(duaknot)){
  cat("knot optimum variabel ke-",i,"\n")
  cat(duaknot[[i]],"\n")
  cat("","\n")
 }
 cat("=====================================================","\n")
 cat("tiga knot optimum untuk masing-masing variabel","\n")
 cat("=====================================================","\n")
 cat("","\n")
 cat(kriteria,"minimum = ",kriteria.3knot,"\n")
 cat("","\n")
 for (i in 1:length(tigaknot)){
  cat("knot optimum variabel ke-",i,"\n")
  cat(tigaknot[[i]],"\n")
  cat("","\n")
 }
 cat("=========================================================","\n")
 cat("hasil kombinasi knot optimum dan ",kriteria," minimum = ","\n")
 cat("=========================================================","\n")
 cat(kriteria,"minimum = ",kriteria.minimum,"\n")
 cat("","\n")
 cat("knot optimum","\n")
 for (i in 1:length(knot.optimum)){
  cat("knot optimum variabel ke-",i,"\n")
  cat(knot.optimum[[i]],"\n")
  cat("","\n")
 }
 cat("","\n")
```

Lampiran 2: Sintaks R Fungsi Utama untuk Memanggil Fungsi Pendukung Lainnya dan Mengeluarkan Output Keseluruhan (lanjutan)

```
cat("=============================================================================================
======================","\n")
 cat("tabel anava","\n")
cat("=============================================================================================
======================","\n")
 cat("Sumber        db      SS            MS        Fhit            P-value             keputusan","\n")
cat("=============================================================================================
======================","\n")
cat("Regresi     ",tabel.anova[1,3],"     ",round(tabel.anova[1,2],3),"     ",as.character(tabel.anova[1,4]),"
",as.character(tabel.anova[1,5]),"      ",as.character(tabel.anova[1,6]),"        ",as.character(tabel.anova[1,7]),"\n")
  cat("Error       ",tabel.anova[2,3],"     ",round(tabel.anova[2,2],3),"     ",as.character(tabel.anova[2,4]),"\n")
  cat("Total       ",tabel.anova[3,3],"     ",round(tabel.anova[3,2],3),"\n")
cat("=============================================================================================
======================","\n")
  cat("","\n")
  cat("","\n")
cat("=============================================================================================
======================","\n")
  cat("parameter beta dan uji individu","\n")
cat("=============================================================================================
======================","\n")
  cat("          beta         standar error       t hitung           p-value             keputusan","\n")
cat("=============================================================================================
======================","\n")
  for (i in 1:nrow(beta)){
    cat("beta",i-1    ," ",beta[i],"       ",inferensi[i,2],"       ",inferensi[i,3],"        ",inferensi[i,4],"
",as.character(inferensi[i,5]),"\n")
  }
cat("=============================================================================================
======================","\n")
  cat("","\n")
  cat("R square =",rsquare,"\n")
  cat("","\n")
  cat("MSE =",mse,"\n")
  cat("","\n")
  cat("","\n")
cat("=============================================================================================
======================","\n")
  cat("Uji kolmogorov-smirnov untuk kenormalan residual ","\n")
cat("=============================================================================================
======================","\n")
  cat("D             P-value             keputusan","\n")
cat("=============================================================================================
======================","\n")
  cat(as.numeric(kenormalan$statistic),"        ",kenormalan$p.value,"       ",keputusan.kenormalan,"\n")
cat("=============================================================================================
======================","\n")
  cat("","\n")
  cat("","\n")
```

Lampiran 2 : Sintaks R Fungsi Utama untuk Memanggil Fungsi Pendukung Lainnya dan Mengeluarkan Output Keseluruhan (lanjutan).

```
 cat("","\n")
cat("============================================================================================
======================","\n")
 cat("uji Gleyjser","\n")
cat("============================================================================================
======================","\n")
 cat("Sumber       db      SS          MS       Fhit          P-value            keputusan","\n")
cat("============================================================================================
======================","\n")
 cat("Regresi     ",gleyjser[1,3],"     ",round(gleyjser[1,2],3),"     ",as.character(gleyjser[1,4]),"
",as.character(gleyjser[1,5]),"     ",as.character(gleyjser[1,6]),"     ",as.character(gleyjser[1,7]),"\n")
 cat("Error       ",gleyjser[2,3],"     ",round(gleyjser[2,2],3),"     ",as.character(gleyjser[2,4]),"\n")
 cat("Total       ",gleyjser[3,3],"     ",round(gleyjser[3,2],3),"\n")
cat("============================================================================================
======================","\n")
 cat("Kesimpulan :",kesimpulan.gly,"\n")
 cat("","\n")
 cat("","\n")
cat("============================================================================================
======================","\n")
 cat("uji Durbin Watson ","\n")
cat("============================================================================================
======================","\n")
 cat("DW          P-value         keputusan","\n")
cat("============================================================================================
======================","\n")
 cat(as.numeric(durbin),"        ",p.value.dw,"       ",keputusan.dw,"\n")
cat("============================================================================================
======================","\n")
 cat("Kesimpulan :",kesimpulan.dw,"\n")
 cat("","\n")
 cat("","\n")
cat("============================================================================================
======================","\n")
 cat("nilai estimasi","\n")
cat("============================================================================================
======================","\n")
 cat("                    y        y topi      ","\n")
cat("============================================================================================
======================","\n")
 for (i in 1:nrow(ytopi)){
   cat("pengamatan ke-",i ,"    ",y[i],"    ",ytopi[i],"\n")
 }
cat("============================================================================================
======================","\n")
 cat("","\n")
 }
```

Lampiran 3 : Sintaks R **"matriktruncated.R"** untuk Membuat Matrix X *Truncated* untuk 1 variabel.

```
truncated<-function(x,orde,k)
  # program membuat matriks x truncated untuk x=1 variabel
  # k dibuat menjadi vektor yang elemennya adalah knot knot yang akan digunakan
  {
  library(pracma)
  orde=1
  x<-as.vector(x)
  k<-as.vector(k)
  d<-length(k)
  trun<-matrix(0,nrow=length(x),ncol=orde+d+1)
  xtrun<-matrix(0,nrow=length(x),ncol=orde+d+1)
  for (j in 1:(orde+1)){
    xtrun[,j]<-x^(j-1)}
    for (r in 1:d){
    for (i in 1:length(x)){
       trun[i,r+orde+1]<-ifelse(x[i]>=k[r],(x[i]-k[r])^(orde),0)
    }
    }
  for (j in (orde+2):(orde+1+d)){
    xtrun[,j]=trun[,j]
  }
  H=xtrun%*%pinv(t(xtrun)%*%xtrun)%*%t(xtrun)
  hasil=list(xtrun=xtrun,H=H)
return(hasil)
}
```

Lampiran 4 : Sintaks R **"matriksx.R"** untuk Membuat Matrix X *Truncated* untuk *p* variabel.

```
mtruncated<-function(x,knot){
  # program membuat matriks x truncated untuk x=p-variabel
  #knot=list dari k untuk masing masing variabel
  library(pracma)
  x=as.matrix(x)
  p=ncol(x)
  xp=vector("list",p)
  for (r in 1:p){
    xp[[r]]=truncated(x[,r],orde,knot[[r]])$xtrun[,-1]
  }
xt=do.call("cbind",xp)
xtruncated=cbind(rep(1,nrow(x)),xt)
H=xtruncated%*%pinv(t(xtruncated)%*%xtruncated)%*%t(xtruncated)
hasil=list(xtruncated=xtruncated,H=H)
return(hasil)
}
```

Lampiran 5 : Sintaks R **"gcvsatuknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 1 Knot.

```
#fungsi mencari knot optimum untuk 1 knot untuk masing masing variabel
gcv1knot<-function(x,y,b){
 #b adalah banyak titik yang dicoba sebagai knot
 source("matriksx.R")
 source("matriktruncated.R")
 x=as.matrix(x)
 y=as.matrix(y)
 gcv<-as.vector(0)
 ubr<-as.vector(0)
 knot=matrix(0,nrow=b,ncol=ncol(x))
 knott=matrix(0,nrow=b-2,ncol=ncol(x))
 knots=vector("list",b)
 kn=vector("list",ncol(x))
 I=diag(1,nrow=nrow(x),ncol=nrow(x))
 cv=c()
 for (j in 1:ncol(x)){
 knot[,j]=seq(min(x[,j]),max(x[,j]),length.out = b)
 knott[,j]=knot[c(-1,-b),j]
 }
 for (i in 1:nrow(knott)){
  knots[[i]]=knott[i,]
 kn=as.list(knots[[i]])
 Ak=mtruncated(x,kn)$H
 SSE=t(y)%*%(I-Ak)%*%y
 dbe=nrow(x)-ncol(matriksX)
 MSE=SSE/dbe
 gcv[i]=((1/nrow(x))*t(y)%*%(I-Ak)%*%y)/(((1/nrow(x))*sum(diag(I-Ak)))^2)
 ubr[i]=(1/nrow(x))*t(Ak%*%y)%*%(I-Ak)^2%*%(Ak%*%y)+(1/nrow(x))*MSE*sum(diag(Ak^2))
 cv[i]=(1/nrow(x))*sum(((y-(Ak%*%y))/(rep(1,nrow(Ak))-diag(Ak)))^2)
  }
 optimum=c(ubr=min(ubr),cv=min(cv),gcv=min(gcv))
 indeks.knotoptimum=c(which.min(ubr),which.min(cv),which.min(gcv))
 knot.optimum=knott[c(indeks.knotoptimum[1],indeks.knotoptimum[2],indeks.knotoptimum[3]),]
 row.names(knot.optimum)=c("knot.optimum.ubr","knot.optimum.cv","knot.optimum.gcv")
 colnames(knot.optimum)=c(paste0("x",1:ncol(x)))
 u1=cbind(knott,ubr)
 sort.u1=u1[order(u1[,6]),]
 urutubr1=sort.u1[1:5,]
 colnames(urutubr1)=c("X1","X2","X3","X4","X5","UBR")
 cat("Berikut tampilan lima nilai UBR terkecil untuk satu knot dan titik-titik knotnya:\n")
 cat("============================================================\n")
 print(urutubr1)
 cat("============================================================\n")
 cat("","\n")
 cat("","\n")
```

Lampiran 5 : Sintaks R **"gcvsatuknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 1 Knot (lanjutan).

```
 c1=cbind(knott,cv)
 sort.c1=c1[order(c1[,6]),]
 urutcv1=sort.c1[1:5,]
 colnames(urutcv1)=c("X1","X2","X3","X4","X5","CV")
 cat("Berikut tampilan lima nilai CV terkecil untuk satu knot dan titik-titik knotnya:\n")
 cat("============================================================\n")
 print(urutcv1)
 cat("============================================================\n")
 cat("","\n")
 cat("","\n")

 hasil<-list(knot=knott,ubr=ubr,cv=cv,gcv=gcv,kriteria.minimum=optimum,knot.optimum=knot.optimum)
 return(hasil)
}
```

Lampiran 6 : Sintaks R **"gcvduaknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 2 Knot.

```
gcv2knot<-function(x,y,b){
 #b adalah banyak titik yang dicoba sebagai knot
 source("matriksx.R")
 source("matriktruncated.R")
 x=as.matrix(x)
 y=as.matrix(y)
 gcv<-as.vector(0)
 ubr<-as.vector(0)
 knot=matrix(0,nrow=b,ncol=ncol(x))
 knott=matrix(0,nrow=b-2,ncol=ncol(x))
 kn=vector("list",ncol(x))
 a=vector("list",ncol(x))
 I=diag(1,nrow=nrow(x),ncol=nrow(x))
 cv=c()
 for (j in 1:ncol(x)){
  knot[,j]=seq(min(x[,j]),max(x[,j]),length.out = b)
  knott[,j]=knot[c(-1,-b),j]
  a[[j]]=t(combn(knott[,j],2))
  }
 for (i in 1:nrow(a[[1]])){
  kn=lapply(a,"[",i,1:2)
  Ak=mtruncated(x,kn)$H
  SSE=t(y)%*%(I-Ak)%*%y
  dbe=nrow(x)-ncol(matriksX)
  MSE=SSE/dbe
  gcv[i]=((1/nrow(x))*t(y)%*%(I-Ak)%*%y)/(((1/nrow(x))*sum(diag(I-Ak)))^2)
  ubr[i]=(1/nrow(x))*t(Ak%*%y)%*%(I-Ak)^2%*%(Ak%*%y)+(1/nrow(x))*MSE*sum(diag(Ak^2))
  cv[i]=(1/nrow(x))*sum(((y-(Ak%*%y))/(rep(1,nrow(Ak))-diag(Ak)))^2)
```

Lampiran 6 : Sintaks R **"gcvduaknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 2 Knot (lanjutan).

```
 }
 optimum=c(ubr=min(ubr),cv=min(cv),gcv=min(gcv))
 indeks.knotoptimum=c(which.min(ubr),which.min(cv),which.min(gcv))
 knot.optimum=lapply(a,"[",indeks.knotoptimum,1:2)
 for (i in 1:length(knot.optimum)){
  row.names(knot.optimum[[i]])=c("knot.optimum.ubr","knot.optimum.cv","knot.optimum.gcv")
  colnames(knot.optimum[[i]])=c(paste0("knot",1:2))
 }
 rr=cbind(1:length(ubr),ubr)
 colnames(rr)<-c("id","ubr")
 urut=rr[order(ubr),]
 urut=round(urut,5)
 cat("Berikut tampilan lima nilai UBR terkecil untuk dua knot dan titik-titik knotnya:\n")
cat("=============================================================================================
============================================","\n")
 cat("            ",paste0("x",1:ncol(x),"            "),"  UBR","\n")
cat("=============================================================================================
============================================","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][1],1))),5),"       "),"    ",urut[,"ubr"][1],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][1],2))),5),"       "),"\n")
 cat("---------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][2],1))),5),"       "),"    ",urut[,"ubr"][2],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][2],2))),5),"       "),"\n")
 cat("---------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][3],1))),5),"       "),"    ",urut[,"ubr"][3],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][3],2))),5),"       "),"\n")
 cat("---------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][4],1))),5),"       "),"    ",urut[,"ubr"][4],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][4],2))),5),"       "),"\n")
 cat("---------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][5],1))),5),"       "),"    ",urut[,"ubr"][5],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][5],2))),5),"       "),"\n")
cat("=============================================================================================
============================================","\n")
 cat("","\n")
 cat("","\n")
 rr1=cbind(1:length(cv),cv)
 colnames(rr1)<-c("id","cv")
 urut1=rr1[order(cv),]
 urut1=round(urut1,5)
 cat("Berikut tampilan lima nilai CV terkecil untuk dua knot dan titik-titik knotnya:\n")
cat("=============================================================================================
============================================","\n")
 cat("            ",paste0("x",1:ncol(x),"            "),"  CV","\n")
cat("=============================================================================================
============================================","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][1],1))),5),"       "),"    ",urut1[,"cv"][1],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][1],2))),5),"       "),"\n")
```

Lampiran 6 : Sintaks R **"gcvduaknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 2 Knot (lanjutan).

```
 cat("-------------------------------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][2],1))),5),"       "),"    ",urut1[,"cv"][2],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][2],2))),5),"       "),"\n")
 cat("-------------------------------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][3],1))),5),"       "),"    ",urut1[,"cv"][3],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][3],2))),5),"       "),"\n")
 cat("-------------------------------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][4],1))),5),"       "),"    ",urut1[,"cv"][4],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][4],2))),5),"       "),"\n")
 cat("-------------------------------------------------------------------------------------------------------------------------------------------","\n")
 cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][5],1))),5),"       "),"    ",urut1[,"cv"][5],"\n")
 cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][5],2))),5),"       "),"\n")

cat("=============================================================================================
==========================================","\n")
 cat("","\n")
 cat("","\n")
 hasil<-list(kriteria.minimum=optimum,knot.optimum=knot.optimum)
 return(hasil)
}
```

Lampiran 7 : Sintaks R **"gcvtigaknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 3 Knot.

```
gcv3knot<-function(x,y,b){
 #b adalah banyak titik yang dicoba sebagai knot
 source("matriksx.R")
 source("matriktruncated.R")
 x=as.matrix(x)
 y=as.matrix(y)
 gcv<-as.vector(0)
 ubr<-as.vector(0)
 knot=matrix(0,nrow=b,ncol=ncol(x))
 knott=matrix(0,nrow=b-2,ncol=ncol(x))
 kn=vector("list",ncol(x))
 a=vector("list",ncol(x))
 I=diag(1,nrow=nrow(x),ncol=nrow(x))
 cv=c()
 for (j in 1:ncol(x)){
  knot[,j]=seq(min(x[,j]),max(x[,j]),length.out = b)
  knott[,j]=knot[c(-1,-b),j]
  a[[j]]=t(combn(knott[,j],3))
 }
 for (i in 1:nrow(a[[1]])){
  kn=lapply(a,"[",i,1:3)
  Ak=mtruncated(x,kn)$H
  SSE=t(y)%*%(I-Ak)%*%y
```

Lampiran 7 : Sintaks R **"gcvtigaknotminimum.Rt"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 3 Knot (lanjutan).

```
dbe=nrow(x)-ncol(matriksX)
   MSE=SSE/dbe
   gcv[i]=((1/nrow(x))*t(y)%*%(I-Ak)%*%y)/(((1/nrow(x))*sum(diag(I-Ak)))^2)
   ubr[i]=(1/nrow(x))*t(Ak%*%y)%*%(I-Ak)^2%*%(Ak%*%y)+(1/nrow(x))*MSE*sum(diag(Ak^2))
   cv[i]=(1/nrow(x))*sum(((y-(Ak%*%y))/(rep(1,nrow(Ak))-diag(Ak)))^2)
   }
  optimum=c(ubr=min(ubr),cv=min(cv),gcv=min(gcv))
  indeks.knotoptimum=c(which.min(ubr),which.min(cv),which.min(gcv))
  knot.optimum=lapply(a,"[",indeks.knotoptimum,1:3)
  for (i in 1:length(knot.optimum)){
   row.names(knot.optimum[[i]])=c("knot.optimum.ubr","knot.optimum.cv","knot.optimum.gcv")
   colnames(knot.optimum[[i]])=c(paste0("knot",1:3))
  }
  rr=cbind(1:length(ubr),ubr)
  colnames(rr)<-c("id","ubr")
  urut=round(rr[order(ubr),],5)
  cat("Berikut tampilan lima nilai UBR terkecil untuk tiga knot dan titik-titik knotnya:\n")

cat("==================================================================================
=========================================================","\n")
  cat("              ",paste0("x",1:ncol(x),"              "),"  UBR","\n")

cat("==================================================================================
=========================================================","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][1],1))),5),"        "),"    ",urut[,"ubr"][1],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][1],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][1],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][2],1))),5),"        "),"    ",urut[,"ubr"][2],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][2],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][2],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][3],1))),5),"        "),"    ",urut[,"ubr"][3],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][3],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][3],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][4],1))),5),"        "),"    ",urut[,"ubr"][4],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][4],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][4],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][5],1))),5),"        "),"    ",urut[,"ubr"][5],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][5],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut[,"id"][5],3))),5),"        "),"\n")

cat("==================================================================================
=========================================================","\n")
  cat("","\n")
  cat("","\n")
```

Lampiran 7 : Sintaks R **"gcvtigaknotminimum.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan 3 Knot (lanjutan).

```
  rr1=cbind(1:length(cv),cv)
  colnames(rr1)<-c("id","cv")
  urut1=round(rr1[order(cv),],5)
  cat("Berikut tampilan lima nilai CV terkecil untuk dua knot dan titik-titik knotnya:\n")

cat("==================================================================================
==========================================================","\n")
  cat("              ",paste0("x",1:ncol(x),"              "),"  CV","\n")

cat("==================================================================================
==========================================================","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][1],1))),5),"        "),"    ",urut1[,"cv"][1],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][1],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][1],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][2],1))),5),"        "),"    ",urut1[,"cv"][2],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][2],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][2],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][3],1))),5),"        "),"    ",urut1[,"cv"][3],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][3],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][3],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][4],1))),5),"        "),"    ",urut1[,"cv"][4],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][4],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][4],3))),5),"        "),"\n")
  cat("-----------------------------------------------------------------------------------------------------------------------------
---","\n")
  cat("knot 1      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][5],1))),5),"        "),"    ",urut1[,"cv"][5],"\n")
  cat("knot 2      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][5],2))),5),"        "),"\n")
  cat("knot 3      ",paste0(round(as.numeric(t(lapply(a,"[",urut1[,"id"][5],3))),5),"        "),"\n")

cat("==================================================================================
==========================================================","\n")
  cat("","\n")
  cat("","\n")
  hasil<-list(kriteria.minimum=optimum,knot.optimum=knot.optimum)
  return(hasil)
}
```

Lampiran 8 : Sintaks R **"kombinasiknotcv.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan Kombinasi Titik Knot Menggunakan Metode CV.

```
kombinasi.knot.cv=function(x,y,b){
 x=as.matrix(x)
 y=as.matrix(y)
 I=diag(1,nrow=nrow(x),ncol=nrow(x))
 cv=c()
 source("matriksx.R")
 source("matriktruncated.R")
 source("gcvsatuknotminimum.R")
 source("gcvduaknotminimum.R")
 source("gcvtigaknotminimum.R")
 k1=gcv1knot(x,y,b)
 knot1=k1$knot.optimum
 knot1.cv=knot1[2,]

 k2=gcv2knot(x,y,b)
 knot2=k2$knot.optimum
 knot2.cv=lapply(knot2,"[",2,1:2)

 k3=gcv3knot(x,y,b)
 knot3=k3$knot.optimum
 knot3.cv=lapply(knot3,"[",2,1:3)

 knott=vector("list",ncol(x))
 kno=vector("list",ncol(x))
 for (i in 1:ncol(x)){
  knott[[i]]=list(knot1.cv[[i]],knot2.cv[[i]],knot3.cv[[i]])
  }
 kom=expand.grid(knott)
 n.knot=matrix(nrow=nrow(kom),ncol=ncol(kom))

 for (i in 1:nrow(kom)){
  kno=lapply(kom,"[[",i)
  Ak=mtruncated(x,kno)$H
  SSE=t(y)%*%(I-Ak)%*%y
  dbe=nrow(x)-ncol(matriksX)
  MSE=SSE/dbe
  cv[i]=(1/nrow(x))*sum(((y-(Ak%*%y))/(rep(1,nrow(Ak))-diag(Ak)))^2)
  for (j in 1:ncol(kom)){
   n.knot[i,j]=length(unlist(kom[i,j]))
  }
  }

 tabel=data.frame(n.knot,cv)
 names(tabel)=c(paste0("x",1:ncol(kom)),"cv")
 tabel=tabel[order(cv),]
 write.csv2(tabel,"kombinasi.csv")
 tabel.baru=(round(tabel[1:5,],4))
 cat("============================================================","\n")
 cat("kombinasi banyak knot untuk masing-masing variabel","\n")
```

Lampiran 8 : Sintaks R **"kombinasiknotcv.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan Kombinasi Titik Knot Menggunakan Metode CV (lanjutan).

```
 cat("============================================================","\n")
 print(tabel.baru,row.names=F)
 cat("============================================================","\n")
 cat("","\n")
 cat("","\n")

 optimum=min(cv)
 indeks.knotoptimum=which.min(cv)
 knot.optimum=lapply(kom,"[[",indeks.knotoptimum)
 hasil<-
list(cv.1knot=k1$kriteria.minimum["cv"],cv.2knot=k2$kriteria.minimum["cv"],cv.3knot=k3$kriteria.minimum["cv"],
satuknot.optimum=knot1.cv,duaknot.optimum=knot2.cv,tigaknot.optimum=knot3.cv,cv.minimum=optimum,knot.
optimum=knot.optimum)
 return(hasil)
}
```

Lampiran 9 : Sintaks R **"kombinasiknotubr.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan Kombinasi Titik Knot Menggunakan Metode UBR.

```
kombinasi.knot.ubr=function(x,y,b){
 x=as.matrix(x)
 y=as.matrix(y)
 ubr<-as.vector(0)
 I=diag(1,nrow=nrow(x),ncol=nrow(x))
 source("matriksx.R")
 source("matriktruncated.R")
 source("gcvsatuknotminimum.R")
 source("gcvduaknotminimum.R")
 source("gcvtigaknotminimum.R")
 k1=gcv1knot(x,y,b)
 knot1=k1$knot.optimum
 knot1.ubr=knot1[1,]

 k2=gcv2knot(x,y,b)
 knot2=k2$knot.optimum
 knot2.ubr=lapply(knot2,"[",1,1:2)

 k3=gcv3knot(x,y,b)
 knot3=k3$knot.optimum
 knot3.ubr=lapply(knot3,"[",1,1:3)

 knott=vector("list",ncol(x))
 kno=vector("list",ncol(x))
  for (i in 1:ncol(x)){
   knott[[i]]=list(knot1.ubr[[i]],knot2.ubr[[i]],knot3.ubr[[i]])
```

Lampiran 9 : Sintaks R **"kombinasiknotubr.R"** untuk Mencari Titik Knot Optimal Model Regresi Nonparametrik Spline *Truncated* dengan Kombinasi Titik Knot Menggunakan Metode UBR (lanjutan).

```
 }
 kom=expand.grid(knott)
 n.knot=matrix(nrow=nrow(kom),ncol=ncol(kom))

 for (i in 1:nrow(kom)){
  kno=lapply(kom,"[[",i)
  Ak=mtruncated(x,kno)$H
  SSE=t(y)%*%(I-Ak)%*%y
  dbe=nrow(x)-ncol(matriksX)
  MSE=SSE/dbe
  ubr[i]=(1/nrow(x))*t(Ak%*%y)%*%(I-Ak)^2%*%(Ak%*%y)+(1/nrow(x))*MSE*sum(diag(Ak^2))
  for (j in 1:ncol(kom)){
   n.knot[i,j]=length(unlist(kom[i,j]))
  }
 }

 tabel=data.frame(n.knot,ubr)
 names(tabel)=c(paste0("x",1:ncol(kom)),"ubr")
 tabel=tabel[order(ubr),]
 write.csv2(tabel,"kombinasi.csv")
 tabel.baru=(round(tabel[1:5,],4))
 cat("=================================================================","\n")
 cat("kombinasi banyak knot untuk masing-masing variabel","\n")
 cat("=================================================================","\n")
 print(tabel.baru,row.names=F)
 cat("=================================================================","\n")
 cat("","\n")
 cat("","\n")
 optimum=min(ubr)
 indeks.knotoptimum=which.min(ubr)
 knot.optimum=lapply(kom,"[[",indeks.knotoptimum)
 hasil<-
list(ubr.1knot=k1$kriteria.minimum["ubr"],ubr.2knot=k2$kriteria.minimum["ubr"],ubr.3knot=k3$kriteria.minimum
["ubr"],satuknot.optimum=knot1.ubr,duaknot.optimum=knot2.ubr,tigaknot.optimum=knot3.ubr,ubr.minimum=op
timum,knot.optimum=knot.optimum)
 return(hasil)
}
```

Lampiran 10 : Sintaks R **"hasilperhitungan.R"** untuk Mencari Nilai Estimasi β dan Uji Signifikansi Parameter

```
hitung<-function(x,y,knot,taraf.alpha){
 # program penaksiran beta dan pengujian hipotesis pada suatu titik knot yang ditentukan dan suatu taraf alpha
 # knot adalah suatu list dari vektor knot maksimal 3 knot untuk masing masing variabel
 library(pracma)
 library(car)
 source("matriksx.R")
 source("matriktruncated.R")
 x=as.matrix(x)
 y=as.matrix(y)
 I=diag(1,nrow=nrow(x),ncol=nrow(x))
 J=matrix(1,nrow=nrow(x),ncol=nrow(x))
 matriksX=mtruncated(x,knot)$xtruncated
 Ak=mtruncated(x,knot)$H
 beta=pinv(t(matriksX)%*%matriksX)%*%t(matriksX)%*%y
 SSE=t(y)%*%(I-Ak)%*%y
 SST=t(y)%*%(I-(1/nrow(x))*J)%*%y
 SSR=SST-SSE
 R.square=as.numeric((SSR/SST)*100)
 dbr=ncol(matriksX)-1
 dbe=nrow(x)-ncol(matriksX)
 dbt=nrow(x)-1
 MSE=SSE/dbe
 MSR=SSR/dbr
 Se=sqrt(MSE*diag(pinv(t(matriksX)%*%matriksX)))
 Fhitung=MSR/MSE
 p.valueF=pf(Fhitung,dbr,dbe,lower.tail=FALSE )
 keputusan=NULL
 keputusan.1=NULL
 if (p.valueF >= taraf.alpha)
 {keputusan.1=c("gagal tolak Hnol")}
 else {keputusan.1=c("tolak hnol")}
 for (i in 1:length(Se)){
  thitung=beta/Se
  p.value=2*pt(abs(thitung),df=nrow(x)-1,lower=FALSE)
  if (p.value[i]>=taraf.alpha) {
   keputusan[i]=c("gagal tolak Hnol")
  } else{
   keputusan[i]=c("tolak Hnol")
  }
 }
 ytopi=Ak%*%y
 b<-
data.frame(Sumber=c("regresi","error","total"),Sum.of.Square=c(round(SSR,4),round(SSE,4),round(SST,4)),db=c(db
r,dbe,dbt),mean.of.square=c(round(MSR,4),round(MSE,4),"-"),Fhitung=c(round(Fhitung,4),"-","-
"),p.value=c(p.valueF,"-","-"),keputusan=c(keputusan.1,"-","-"))
 inferensi=data.frame(beta=beta,standar.error=Se,t.hitung=thitung,p.value=p.value,keputusan=keputusan)
 hasil<-list(ytopi=ytopi,beta=beta,tabel.anova=b,inferensi=inferensi,R.square=R.square,MSE=MSE)
 return(hasil)
}
```

Lampiran 11 : Sintaks R **"plotgab.R"** untuk Membuat Plot Gabungan Antara Metode CV dan UBR.

```
#plot gabungan
b=35
ubr=kombinasi.knot.ubr(x,y,b)
ubr.knot.optimum=ubr$knot.optimum
cv=kombinasi.knot.cv(x,y,b)
cv.knot.optimum=cv$knot.optimum
ytopi.ubr=hitung(x,y,ubr.knot.optimum,taraf.alpha)$ytopi
ytopi.cv=hitung(x,y,cv.knot.optimum,taraf.alpha)$ytopi
plot(y,xlab="pengamatan",ylab="y.dan.ytopi",type="b",pch=19,col="red3")
lines(ytopi.ubr,col="cyan3",type="b",pch=18,lty=2)
lines(ytopi.cv,col="cyan4",type="b",pch=18,lty=2)
legend("topleft", legend=c("y", "y topi ubr","y topi cv"),inset=.02,
      col=c("red3", "cyan3", "cyan4"), lty=1:3, cex=0.8)
```

Lampiran 12 : Sintaks Simulasi

```
simulasi=function(b=30,varian=0.025,n=100,replikasi=20){
  source("simulasi_cv_ubr (fs panggil).R")
  mse.ubr=c()
  mse.cv=c()
  rsq.ubr=c()
  rsq.cv=c()
  knot.cv.var1=c()
  knot.cv.var2=c()
  knot.ubr.var1=c()
  knot.ubr.var2=c()
  e=c()
  x1s=c()
  x2s=c()
  for (l in 1:replikasi){
    cat("replikasi ke-",l,"\n")
    for(i in 1:n){
      e[i]=rnorm(1,mean=0,sd=sqrt(varian))
      x1s[i]=runif(1,0,1)
      x2s[i]=runif(1,0,1)
    }
    xsim=cbind(x1s,x2s)
    ysim1=4+2*x1s+3.5*ifelse(x1s<=0.5,0,x1s-0.5)+1.5*x2s+4*ifelse(x2s<=0.6,0,x2s-0.6)
    ysim=ysim1+e
    a=simulasi.spline.multivariabel.ubr.cv2(xsim,ysim,b,0.05)
    mse.ubr[l]=as.numeric(a$mse.ubr)
    mse.cv[l]=as.numeric(a$mse.cv)
    rsq.ubr[l]=as.numeric(a$Rsq.ubr)
    rsq.cv[l]=as.numeric(a$Rsq.cv)
```

Lampiran 12 : Sintaks Simulasi (lanjutan)

```
    knot.cv.var1[l]=as.numeric(a$knot.cv[1])
    knot.cv.var2[l]=as.numeric(a$knot.cv[2])
    knot.ubr.var1[l]=as.numeric(a$knot.ubr[1])
    knot.ubr.var2[l]=as.numeric(a$knot.ubr[2])
  }

tabel.knot=data.frame(knot.cv.var1=knot.cv.var1,knot.cv.var2=knot.cv.var2,knot.ubr.var1=knot.ubr.var1,knot.ubr.
var2=knot.ubr.var2)
  tabel.mse=data.frame(mse.cv=mse.cv,mse.ubr=mse.ubr)
  tabel.rsquare=data.frame(rsq.cv=rsq.cv,rsq.ubr=rsq.ubr)
  row.names(tabel.knot)=c(paste0("replikasi",1:replikasi))
  row.names(tabel.mse)=c(paste0("replikasi",1:replikasi))
  row.names(tabel.rsquare)=c(paste0("replikasi",1:replikasi))
  hasil=list(tabel.knot=tabel.knot,tabel.mse=tabel.mse,tabel.rsquare=tabel.rsquare)
  return(hasil)
}
```

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika Fakultas Matematika, Komputasi, dan Sains Data (FMKSD) ITS:

Nama : Alvita Rachma Devi

NRP : 06211650010005

Program Studi : Magister Statistika / ~~Doktor Ilmu Statistik~~

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

Sumber : BPS

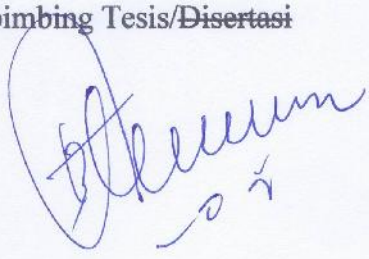
Keterangan : Profil Ketenagakerjaan Jawa Tengah Hasil Sakernas Agustus 2012, Provinsi Jawa Tengah dalam Angka 2017, Statistik Pendidikan Jawa Tengah 2015, website BPS Jawa Tengah

Surat pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, 21 Desember 2017

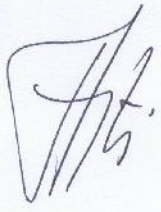
Mengetahui

Pembimbing Tesis/~~Disertasi~~



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

Mahasiswa



Alvita Rachma Devi

*(coret yang tidak perlu)

** Mohon dituliskan data primer apabila datanya dikumpulkan sendiri

BIOGRAFI PENULIS

Alvita Rachma Devi, lahir di Semarang, Jawa Tengah pada tanggal 7 Januari 1992. Penulis merupakan putri pertama dari pasangan Bapak Drs. Sudargo dan Ibu Dra. Suparti. Penulis memulai jenjang pendidikan formal di SD Negeri Gayamsari 02-05 Semarang pada tahun 1998. Kemudian penulis melanjutkan pendidikan di SMP Negeri 2 Semarang tahun 2004 dan SMA Negeri 3 Semarang pada tahun 2007. Pada tahun 2010, penulis melanjutkan pendidikan Sarjana (S1) di Universitas Diponegoro Semarang pada Jurusan Statistika dan memperoleh gelar sarjana sains pada bulan April 2014. Penulis melanjutkan studi Pasca Sarjana (S2) di Institut Teknologi Sepuluh Nopember Surabaya Jurusan Statistika pada tahun 2016. Segala kritik dan saran yang berkaitan dengan tesis ini dapat disampaikan kepada penulis melalui email : alvita.erdhe@gmail.com.