**ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN *SPLINE TRUNCATED***

**( Skripsi )**

**Oleh**

**M. RISKI MULTAZAM**

**JURUSAN MATEMATIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS LAMPUNG**
**BANDAR LAMPUNG**
**2019**

## ABSTRACT

## THE ESTIMATION OF NONPARAMETRIC REGRESSION CURVES USING SPLINE TRUNCATED

**By**

**M. Riski Multazam**

The purpose of this research is to find out whether the spline truncated method depends on the distribution of error. Simulation data from trigonometry function with different error distribution are used. The result of this research shows that the distribution of error affected the estimation of spline truncated function. The spline function is estimated better when error is normally distributed than nonnormally distributed.

*Keywords: Nonparametric Regression, Spline Truncated, Error.*

## ABSTRAK

**ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN *SPLINE TRUNCATED***

**Oleh**

**M. Riski Multazam**

Tujuan dari penelitian ini adalah untuk mengetahui apakah metode *spline truncated* bergantung pada distribusi galat. Simulasi data dari fungsi trigonometri dengan menggunakan distribusi galat yang berbeda. Hasil dari penelitian ini menunjukkan bahwa distribusi galat memengaruhi estimasi *spline truncated*. Fungsi *spline* diperkirakan lebih baik ketika galat berdistribusi normal dibandingkan galat berdistribusi tidak normal.

Kata kunci: Regresi Nonparametrik, *Spline Truncated*, Galat.

# ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN *SPLINE TRUNCATED*

**Oleh**

*M. Riski Multazam*

**Skripsi**
**Sebagai salah satu syarat untuk mencapai gelar**
**SARJANA SAINS**

**Pada**
**Jurusan Matematika**
**Fakultas Matematika dan Ilmu Pengetahuan Alam**

**JURUSAN MATEMATIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS LAMPUNG**
**BANDAR LAMPUNG**
**2019**

| | |
|---|---|
| Judul Skripsi | : **ESTIMASI KURVA REGRESI NONPARAMETRIK MENGGUNAKAN *SPLINE TRUNCATED*** |
| Nama Mahasiswa | : *M. Riski Multazam* |
| Nomor Pokok Mahasiswa | : 1517031003 |
| Jurusan | : Matematika |
| Fakultas | : Matematika dan Ilmu Pengetahuan Alam |

**MENYETUJUI**

1. Komisi Pembimbing

| | |
|---|---|
| **Ir. Netti Herawati, M.Sc., Ph.D.** | **Dr. Muslim Ansori, S.Si., M.Si.** |
| NIP. 196501251990032001 | NIP. 197202271998021001 |

2. Mengetahui

Ketua Jurusan Matematika

**Prof. Dra. Wamiliana, M.A., Ph.D.**
NIP. 19631108 198902 2 001

## MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ........................

Sekretaris : **Dr. Muslim Ansori, S.Si., M.Si.** ........................

Penguji
Bukan Pembimbing : **Drs. Nusyirwan, M.Si.** ........................

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

**Drs. Suratman, M.Sc.**
NIP. 196406041990031002

Tanggal Lulus Ujian Skripsi : **24 Mei 2019**

**PERNYATAAN SKRIPSI MAHASISWA**

Saya yang bertanda tangan dibawah ini:

| | |
|---|---|
| Nama | : M. Riski Multazam |
| Nomor Pokok Mahasiswa | : 1517031003 |
| Judul | : Estimasi Kurva Regresi Nonparametrik Menggunakan *Spline Truncated* |
| Jurusan | : Matematika |

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri, bukan hasil orang lain, dan semua hasil tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 24 Mei 2019

Penulis

**M. Riski Multazam**
**NPM. 1517031003**

## RIWAYAT HIDUP

Penulis dilahirkan di Kasui, Kab. Way Kanan pada 29 Mei 1997 sebagai anak keempat dari Bapak H. A. Gani, S.Ag. dan Ibu Sri Lita W.P.

Pendidikan Sekolah Dasar (SD) diselesaikan di SDN 1 Kasui Pasar, Kab. Way Kanan pada tahun 2009. Kemudian, penulis menyelesaikan Sekolah Menengah Pertama (SMP) di SMPN 1 Kasui, Kab. Way Kanan pada tahun 2012. Pada tahun 2015, penulis menyelesaikan Sekolah Menengah Atas (SMA) di SMAN 1 Kasui, Kab. Way Kanan.

Tahun 2015, penulis terdaftar sebagai mahasiswa Jurusan Matematika FMIPA Unila melalui jalur SNMPTN. Selama menjadi mahasiswa, penulis cukup aktif menjadi staff ahli BEM UKBM Unila. Pada awal tahun 2018, penulis melaksanakan Kuliah Praktik di Badan Pusat Statistik (BPS) Kota Metro dan melaksanakan KKN di Desa Itik Renday, Kecamatan Melinting, Kabupaten Lampung Timur.

## KATA INSPIRASI

*“Dan janganlah kamu berputus asa dari rahmat Allah. Sesungguhnya tiada terputus dari rahmat Allah melainkan orang-orang yang kufur’’*

(Q.S. Yusuf: 87)

*“Allah tidak akan membebani seseoarang melainkan sesuai dengan kadar kesanggupannya’’*

(Q.S. Al-Baqarah: 286)

*‘‘Sesungguhnya bersama kesulitan ada kemudahan’’*

(Q.S. Al-Insyirah: 6)

*‘‘Be yourself, but always be your better self’’*

(Karl Maeser)

*‘‘Berbicara tentang akan menjadi apa aku nanti*
*Aku tidak akan mengatakannya*
*Aku akan menunjukkannya’’*

(M. Riski Multazam**)**

## *PERSEMBAHAN*

*Dengan mengucap Alhamdulillah atas berkat dan rahmat Allah SWT kupersembahkan karya kecilku ini untuk:*

*Ayah dan ibuku tercinta yang telah mencurahkan seluruh hidupnya untuk kebahagiaanku dan tak berhenti untuk selalu mendoakanku.*

*Ketiga kakak dan adik kecilku serta seluruh keluarga dekat yang telah mendukung.*

*Dosen pembimbing dan dosen penguji yang telah berjasa dan selalu memberikan motivasi kepada penulis.*

*Orang terdekatku, sahabat-sahabatku, dan Almamaterku Universitas Lampung.*

*Terima Kasih.*

## SANWACANA

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya kepada penulis sehingga skripsi ini dapat diselesaikan dengan sebaik-baiknya. Shalawat dan salam semoga selalu tercurah kepada Nabi Muhammad SAW sebagai suri tauladan bagi kita.

Skripsi dengan judul "Estimasi Kurva Regresi Nonparametrik Menggunakan *Spline Truncated*" adalah salah satu syarat untuk memperoleh gelar sarjana Matematika di Universitas Lampung.

Dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku Pembimbing Akademik sekaligus Pembimbing Utama atas kesediaannya untuk memberikan bimbingan, saran, dan kritik dalam proses penyelesaian skripsi ini;
2. Bapak Dr. Muslim Ansori, S.Si., M.Si., selaku Pembimbing Kedua atas kesediaannya untuk memberikan bimbingan, saran, dan kritik dalam proses penyelesaian skripsi ini;
3. Bapak Drs. Nusyirwan, M.Si., selaku Penguji pada ujian skripsi. Terima kasih untuk masukan dan saran-saran pada seminar proposal penelitian terdahulu;
4. Ibu Prof. Dra. Wamiliana, MA, Ph.D., selaku Ketua Jurusan Matematika.
5. Bapak Drs. Suratman, M.Sc., selaku Dekan FMIPA Unila.

6. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Orang tua yang telah mencurahkan seluruh hidup dan kasih sayangnya sehingga penulis dapat bertahan hingga saat ini;
8. Kakak-kakak dan adik (Ryan Dewi P., M. Emir Musaddad, Deby Putri R., dan M. Fatih Ramadhan) yang sudah menjadi saudara terbaik bagi penulis;
9. Ida, Lisa, Robin, Winda, Sulis, Ratih, Sonia, Mega, Reza dan teman-teman alumni SMA N 1 Kasui yang telah memberikan semangat bagi penulis;
10. Ario, Dony, Rahmad, Siska, Putri C.B., Nurhayati, Windi, Irmawati, Yulia, May, Riana, Aimila, Indah, dan Fransiska yang telah memberikan semangat dan motivasi bagi penulis selama masa perkuliahan;
11. Tira, Riswanti, Neli, Meilinda, Irma Ningsih, Dita, dan Wahyu yang sudah saling membantu dan memberikan motivasi dalam menyelesaikan skripsi;
12. Teman-teman KKN desa Itik Renday, Lampung Timur (Cynthia, Citra, Mira, Alesia, Alan, dan Febri) yang telah memberikan dukungan;
13. Teman-teman matematika 2015;
14. Seluruh pihak yang telah membantu penulis yang tidak dapat disebutkan satu persatu, atas peran dan dukungannya dalam menyusun skripsi ini.

Akhir kata, penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan akan tetapi semoga skripsi yang sederhana ini dapat berguna dan bermanfaat bagi kita semua.

Bandar Lampung, Mei 2019
Penulis

**M. Riski Multazam**

## DAFTAR ISI

# DAFTAR TABEL

# DAFTAR GAMBAR

# I. PENDAHULUAN

## 1.1. Latar Belakang dan Masalah

Analisis regresi merupakan metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon (*dependen*) dengan variabel prediktor (*independen*). Pendekatan regresi dapat dilakukan dengan tiga pendekatan antara lain pendekatan parametrik, semiparametrik, dan nonparametrik. Pendekatan parametrik dilakukan apabila bentuk kurva diketahui sedangkan pendekatan nonparametrik dilakukan apabila bentuk kurva tidak diketahui. Pendekatan semiparametrik dilakukan apabila sebagian bentuk kurva diketahui dan sebagian tidak diketahui. Pendekatan parametrik sering dilakukan dikarenakan kemudahan dalam estimasinya, tetapi pendekatan parametrik terikat asumsi-asumsi yang sangat kaku dan ketat, bentuk kurva diketahui dan residual berdistribusi normal, homoskedasitas, serta tidak terjadi korelasi antar residual. Jika suatu permasalahan dimodelkan menggunakan pendekatan parametrik namun asumsi-asumsinya tidak terpenuhi maka model tersebut akan menjadi bias untuk digunakan.

Sementara itu, pendekatan nonparametrik tidak memerlukan asumsi-asumsi yang terikat seperti pendekatan parametrik. Pendekatan nonparametrik memiliki fleksibelitas yang tinggi karena tidak mengasumsikan bentuk kurva regresi.

Kurva diharapkan dapat mengestimasi sendiri bentuk kurva berdasarkan sebaran data dengan memerhatikan metode kemulusan (*smooth*) kurva tersebut. Ada beberapa metode pemulusan yang dapat dilakukan pada pendekatan nonparametrik yaitu deret *fourier*, kernel, dan *spline*. Diantara metode-metode pendekatan nonparametrik tersebut, Budiantara (2009) menyebutkan bahwa *spline* merupakan metode yang mempunyai interpretasi statistik dan interpretasi visual yang sangat khusus dan sangat baik. *Spline* mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik/turun yang tajam dengan bantuan titik knot. Titik knot merupakan titik perpaduan bersama dimana terjadi pola perubahan perilaku kurva pada interval yang berbeda. *Spline* sangat dipengaruhi oleh parameter pemulus ($\lambda$) yang optimal untuk menduga bentuk kurva.

Pemilihan $\lambda$ optimal pada hakekatnya merupakan pemilihan lokasi dan jumlah titik knot. Akibatnya, pemilihan $\lambda$ optimal merupakan hal yang sangat penting dalam pendekatan nonparametrik. Salah satu metode pemilihan jumlah titik knot optimal adalah menggunakan metode *Generalized Cross Validation* (GCV).

Dalam penelitian ini akan membahas tentang estimasi kurva regresi nonparametrik menggunakan *spline truncated* pada beberapa kondisi galat yang berbeda berdasarkan nilai GCV.

## 1.2. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk mengetahui apakah estimasi *spline truncated* bergantung pada galat dan kondisi galat seperti apa yang memberikan performa baik dalam estimasi pada *spline truncated*.

### 1.3. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

1. Menambah pengetahuan tentang regresi nonparametrik *spline truncated*.
2. Menambah pengetahuan penentuan jumlah titik knot optimal regresi nonparametrik *spline truncated* berdasarkan GCV.
3. Memperoleh kurva penduga *spline truncated* terbaik berdasarkan nilai $R^2$.

# II. TINJAUAN PUSTAKA

## 2.1. Analisis Regresi

Analisis regresi merupakan suatu analisis statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan variabel respon. Secara umum regresi memiliki bentuk fungsi sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i \ , \quad i = 1, 2, \ldots, n. \qquad (2.1)$$

Dengan $y_i$ adalah variabel respon sedangkan fungsi $f(x_i)$ merupakan kurva regresi dengan $x_i$ sebagai variabel prediktor dan $\varepsilon_i$ adalah galat yang diasumsikan berdistribusi normal (0,$\sigma^2$) (Drapper & Smith, 1992).

## 2.2. Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel penjelas dan variabel respon yang tidak diketahui kurva regresinya. Pendekatan nonparametrik memiliki fleksibelitas yang tinggi karena tidak mengasumsikan bentuk kurva regresi. Dalam hal ini, kurva diharapkan dapat mengestimasi sendiri bentuk kurva berdasarkan sebaran data dengan memerhatikan sifat kemulusan (*smooth*) kurva tersebut.

Secara umum model regresi nonparametrik adalah :

$$y_i = f(x_i) + \varepsilon_i \ , \quad i = 1, 2, \ldots, n. \qquad (2.2)$$

Dengan $y_i$ adalah variabel respon dan $\varepsilon_i$ merupakan galat yang berdistribusi normal, mean nol dan varian $\sigma^2$ (0,$\sigma^2$) (Wahba, 1990). Dalam hal ini, $f(x_i)$ hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, dimana pemilihan ruang fungsi tersebut biasanya termotivasi oleh sifat kemulusan (*smooth*) yang dimiliki oleh fungsi $f(x_i)$ tersebut (Herawati, 2011).

## 2.3. Regresi *Spline Truncated*

Hardle (1990) dan Wahba (1990) menyarankan penggunaan regresi *spline* sebagai alternatif pendekatan nonparametrik. *Spline* merupakan potongan (*truncated*) polinomial tersegmen yang kontinu dan memiliki sifat fleksibilitas, sehingga mampu mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif mulus. Titik knot adalah titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda. Titik knot diambil pada selang $a < K_k < b$ , dimana $a$ adalah nilai minimum dan $b$ adalah nilai maksimum suatu data.

Wu & Zang (2006), menuliskan secara umum fungsi *spline* berorde m adalah sembarang fungsi yang dapat ditulis dalam bentuk sebagai berikut :

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

dengan fungsi *truncated*:

$$(x-K_k)_+^{m-1} = \begin{cases}(x-K_k)^{m-1}, & x \geq K \\ 0 & , x < K\end{cases}$$

$\beta_j$ dan $\beta_{m-1+k}$ adalah parameter , $\lambda = \{K_1, K_2, \dots, K_r\}$ adalah titik knot, $f(x)$ adalah fungsi regresi, dan $x$ adalah variabel prediktor, serta $(x-K_k)_+^{m-1}$ merupakan fungsi potongan polinomial tersegmen yang kontinu.

Apabila diberikan model nonparametrik :

$$y_i = f(x_i) + \varepsilon_i, \qquad \text{i = 1, 2, …, n.}$$

dimana adalah $\varepsilon_i$ galat acak dengan mean nol dan varians $\sigma^2$ . Maka, bentuk umum regresi *spline* keluarga polinomial *truncated* orde ke-$m$ dengan satu prediktor adalah

$$y_i = \sum_{j=0}^{m-1} \beta_j x_i^j + \sum_{k=1}^{r} \beta_{m-1+k}(x_i - K_k)_+^{m-1} + \varepsilon_i. \qquad (2.4)$$

Dimana $\beta$ adalah parameter, m adalah orde dengan j=0,1,2,…,m-1, dan K adalah titik knot dengan k=1,2,3,…,r dan i = 1, 2, …, n adalah banyaknya pengamatan.

## 2.4. Estimasi Parameter Regresi Nonparametrik *Spline Truncated*

Hidayat (2017) melakukan suatu kajian dengan kurva regresi $f$ dihampiri dengan fungsi *spline* $f$ dengan titik knot K untuk mendapatkan estimator *spline*. Dalam bentuk matriks disajikan sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Apabila model regresi *spline* disajikan dalam bentuk matriks, diperoleh:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \dots & x_1^{m-1} & (x_1 - K_1)_+^{m-1} & (x_1 - K_2)_+^{m-1} & \dots & (x_1 - K_r)_+^{m-1} \\ 1 & x_2 & \dots & x_2^{m-1} & (x_2 - K_1)_+^{m-1} & (x_2 - K_2)_+^{m-1} & \dots & (x_2 - K_r)_+^{m-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{m-1} & (x_n - K_1)_+^{m-1} & (x_n - K_2)_+^{m-1} & \dots & (x_n - K_r)_+^{m-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{m-1} \\ \beta_{(m-1)+1} \\ \beta_{(m-1)+2} \\ \vdots \\ \beta_{(m-1)+r} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

atau dapat ditulis dengan

$$\tilde{\boldsymbol{y}} = \boldsymbol{X}[K_1, K_2, \dots, K_r]\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}} \tag{2.5}$$

Selanjutnya, estimasi parameter $\tilde{\boldsymbol{\beta}} = [\beta_0\ \beta_1 \dots \beta_m\ \beta_{m+1}\ \beta_{m+2} \dots\ \beta_{m+r}]^T$ diperoleh melalui metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat galat sebagai berikut:

$$\min_{\tilde{\beta} \in R^{m+1+r}} \{\tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}}\} = \min_{\tilde{\beta} \in R^{m+1+r}} \left( \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}^T \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \right)$$

$$\min_{\tilde{\beta} \in R^{m+1+r}} \{\tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}}\} = \min_{\tilde{\beta} \in R^{m+1+r}} \{[[\tilde{\boldsymbol{y}} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\tilde{\boldsymbol{\beta}}]^T [[\tilde{\boldsymbol{y}} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\tilde{\boldsymbol{\beta}}]]\}$$

Dengan penyajian matriks diberikan sebagai berikut:

$$\begin{aligned} \textstyle\sum_{i=1}^{n} \varepsilon_i^2 &= \tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}} \\ &= ([\tilde{\boldsymbol{y}} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\tilde{\boldsymbol{\beta}}]^T [[\tilde{\boldsymbol{y}} - \boldsymbol{X}[K_1, K_2, \dots, K_r]\tilde{\boldsymbol{\beta}}]) \\ &= ([\tilde{\boldsymbol{y}} - \boldsymbol{X}[\boldsymbol{\lambda}]\tilde{\boldsymbol{\beta}}]^T [\tilde{\boldsymbol{y}} - \boldsymbol{X}[\lambda]\tilde{\boldsymbol{\beta}}\ ]) \\ &= \tilde{\boldsymbol{y}}^T \tilde{\boldsymbol{y}} - 2\tilde{\boldsymbol{\beta}}^T \boldsymbol{X}^T[\lambda]\tilde{\boldsymbol{y}} + \tilde{\boldsymbol{\beta}}^T \boldsymbol{X}^T[\boldsymbol{\lambda}]\boldsymbol{X}[\boldsymbol{\lambda}]\tilde{\boldsymbol{\beta}} \end{aligned} \tag{2.6}$$

Kemudian, persamaan (2.8) diturunkan terhadap vektor $\tilde{\beta}^T$ dan disamakan dengan nol, maka diperoleh:

$$\frac{\partial(\tilde{\boldsymbol{\varepsilon}}^T \tilde{\boldsymbol{\varepsilon}})}{\partial \tilde{\boldsymbol{\beta}}^T} = \frac{\partial(\tilde{\boldsymbol{y}}^T \tilde{\boldsymbol{y}} - 2\tilde{\boldsymbol{\beta}}^T \boldsymbol{X}^T[\boldsymbol{\lambda}]\tilde{\boldsymbol{y}} + \tilde{\boldsymbol{\beta}}^T \boldsymbol{X}^T[\boldsymbol{\lambda}]\boldsymbol{X}[\boldsymbol{\lambda}]\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^T}$$

$$\hat{\tilde{\boldsymbol{\beta}}} = (\boldsymbol{X}^{T}[\boldsymbol{\lambda}]\boldsymbol{X}[\boldsymbol{\lambda}])^{-1}\boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}]\tilde{\boldsymbol{y}} \tag{2.7}$$

dengan,

$$\boldsymbol{X}[\boldsymbol{\lambda}] = \boldsymbol{X}[K_1, K_2, \dots, K_r]$$

$$= \begin{bmatrix} 1 & x_1 & \dots & x_1^m & (x_1 - K_1)_+^m & (x_1 - K_2)_+^m & \dots & (x_1 - K_r)_+^m \\ 1 & x_2 & \dots & x_2^m & (x_2 - K_1)_+^m & (x_2 - K_2)_+^m & \dots & (x_2 - K_r)_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^m & (x_n - K_1)_+^m & (x_n - K_2)_+^m & \dots & (x_n - K_r)_+^m \end{bmatrix}$$

## 2.5. Estimasi Kurva Regresi Nonparametrik *Spline Truncated*

Setelah didapatkan estimasi parameter seperti pada persamaan (2.7), maka estimasi titik kurva regresi nonparametrik *spline truncated* dapat diperoleh dengan mensubstitusikan persamaan (2.7) ke dalam persamaan (2.8). Akibatnya, estimasi untuk kurva regresi *spline truncated* dengan knot λ diberikan oleh:

$$\hat{\boldsymbol{y}}[\lambda](x_i) = \hat{\tilde{\boldsymbol{f}}}[\lambda](x_i) = \boldsymbol{X}[\boldsymbol{\lambda}]\hat{\tilde{\boldsymbol{\beta}}} \tag{2.8}$$

$$\hat{\tilde{\boldsymbol{f}}}[\lambda](x_i) = \boldsymbol{X}[\boldsymbol{\lambda}](\boldsymbol{X}^{\boldsymbol{T}}[\boldsymbol{\lambda}]\boldsymbol{X}[\boldsymbol{\lambda}])^{-1}\boldsymbol{X}^{T}[\boldsymbol{\lambda}]\tilde{\boldsymbol{y}}$$

Sehingga diperoleh:

$$\hat{\tilde{\boldsymbol{f}}}[\lambda](x_i) = \boldsymbol{A}[\boldsymbol{\lambda}]\tilde{\boldsymbol{y}} \tag{2.9}$$

dengan, $\boldsymbol{A}[\boldsymbol{\lambda}] = \boldsymbol{X}[\boldsymbol{\lambda}](\boldsymbol{X}^{T}[\boldsymbol{\lambda}]\boldsymbol{X}[\boldsymbol{\lambda}])^{-1}\boldsymbol{X}^{T}[\boldsymbol{\lambda}]$ dan $\boldsymbol{X}[\boldsymbol{\lambda}]$ adalah matriks dari model yang bergantung pada titik knot dan $\lambda = [K_1, K_2, \dots, K_r]^T$ merupakan titik knot.

## 2.6. Pemilihan Banyaknya Titik Knot Optimal

Menurut Eubank (1988), pemilihan $\lambda$ optimal pada hakekatnya merupakan pemilihan lokasi dan banyaknya titik knot. Dalam regresi nonparametrik *spline* sangat penting untuk menentukan banyaknya titik knot optimal, Wan (2000) menyebutkan bahwa strategi untuk memilih banyaknya titik knot dengan menggunakan titik knot yang relatif sedikit. Menurut Lee (2005) alasan atau pertimbangan tersebut berkenaan dengan pemilihan model secara statistika dimana pemilihan titik knot haruslah yang terbaik dan lebih mengarah pada kesederhanaan (*parsimony*) model . Jika diperoleh titik knot optimal maka diperoleh fungsi *spline* terbaik. Budiantara (2006) menyebutkan metode yang dapat digunakan untuk memilih jumlah titik knot optimal adalah dengan menggunakan metode *Generalized Cross Validation* (GCV).

Menurut Wu & Zang (2006), metode GCV dapat dituliskan sebagai berikut

$$\text{GCV}[\lambda] = \frac{n^{-1}\sum_i^n (y_i - \hat{y}[\lambda](x_i))^2}{(n^{-1}\, trace[\boldsymbol{I} - \boldsymbol{A}[\boldsymbol{\lambda}]])^2} \tag{2.10}$$

dengan $\boldsymbol{A}[\boldsymbol{\lambda}] = \boldsymbol{X}[\boldsymbol{\lambda}](\boldsymbol{X}^T[\boldsymbol{\lambda}]\boldsymbol{X}[\boldsymbol{\lambda}])^{-1}\boldsymbol{X}^T[\boldsymbol{\lambda}]$. Selanjutnya, titik knot optimal dipilih dengan cara meminimumkan nilai GCV$[\lambda]$.

## 2.7. Ukuran Kebaikan Model dari Titik Knot Optimal

Salah satu tujuan analisis regresi adalah mendapatkan model terbaik yang mampu menjelaskan hubungan antara variabel prediktor dengan variabel respon. Kebaikan suatu penduga dapat dilihat dari tingkat kesalahannya. Semakin kecil tingkat kesalahan suatu pendugaan maka semakin baik estimasinya.

Menurut Chatterjee (2006), kriteria untuk mentukan estimator terbaik dalam model regresi adalah nilai *Mean Square Error* (MSE) dan nilai koefisien determinasi *R-Square* ($R^2$). Gujarati (2003) menyatakan besaran nilai $R^2$ tidak pernah negatif dan batasannya adalah $0 \leq R^2 \leq 1$. Jika nilai $R^2$ semakin besar atau mendekati nilai satu maka semakin baik pula model yang didapatkan.

MSE didefinisikan sebagai berikut:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \quad (2.11)$$

MSE sangat baik dalam memberikan gambaran terhadap seberapa konsisten model yang dibangun. Dengan meminimalkan nilai MSE, berarti meminimalkan varian model. Model yang memiliki varian kecil mampu memberikan hasil yang relatif lebih konsisten untuk seluruh data input dibandingkan dengan model dengan varian besar (MSE besar).

Sedangkan koefisien determinasi didefinisikan sebagai berikut:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \quad (2.12)$$

Dimana, $y_i$ adalah data variabel respon ke-i, $\bar{y}$ adalah mean data variabel respon, sedangkan $\hat{y}_i$ adalah nilai hasil estimasi variabel respon ke-i. *Sum of Square Regression (SSR)* adalah jumlah kuadrat simpangan hasil dugaan terhadap rata-rata variabel respon. Sedangkan *Sum of Square Total (SST )* adalah jumlah kuadrat simpangan variabel respon. SSR berfungsi untuk mengukur kualitas variabel prediktor sebagai prediktor variabel respon. Sehingga, koefisien determinasi dapat diartikan sebagai proporsi keragaman total variabel respon yang diukur oleh variabel prediktor.

Ghazali (2006) menjelaskan masalah yang terjadi jika melakukan pengujian dengan menggunakan $R^2$ adalah jika variabel bebasnya lebih dari satu maka nilai $R^2$ akan bertambah besar. Pengujian dengan *adjusted R-Square* secara obyektif melihat pengaruh penambahan variabel bebas, apakah variabel tersebut mampu memperkuat variasi penjelasan variabel terikat.

Adapun perhitungan nilai *adjusted R-Square* adalah sebagai berikut:

$$\text{adjusted R-Square} = 1 - (1 - R^2)\left(\frac{n-1}{n-p}\right) \tag{2.13}$$

dengan n adalah banyaknya data observasi dan p banyaknya variabel independen.

# III. METODOLOGI PENELITIAN

## 3.1. Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2018/2019 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

## 3.2. Data Penelitian

Data penelitian berasal dari simulasi dengan membangkitkan data variabel prediktor $(x)$ yang berdistribusi *Uniform* [0,3] sebanyak 100 pengamatan. Data pada selang [0,3] dibangkitkan agar diperoleh sebaran data yang dapat menggambarkan bentuk fungsi sinus maupun cosinus secara jelas. Sementara itu, data variabel respon (y) diperoleh dari persamaan:

1. $y = \sin(4x) + \varepsilon$
2. $y = \cos(4x) + \varepsilon$.

Pada penelitian ini, galat yang dibangkitkan adalah berdistribusi normal dan tidak normal (mengandung pencilan) agar diperoleh informasi apakah *spline truncated* baik dalam mengestimasi data pada kondisi galat tersebut.

Dengan $\varepsilon_i$ yang berdistribusi:

a. $\varepsilon \sim N(0{,}0.3)$

b. $\varepsilon \sim N(0{,}1)$

c. $\varepsilon \sim N(0{,}1.3)$

d. $\varepsilon \sim 0.95N(0{,}1) + 0.05N(5{,}1)$

e. $\varepsilon \sim 0.9N(0{,}1) + 0.1N(5{,}1)$.

### 3.3. Metode Penelitian

Dalam penelitian ini akan dilakukan perbandingan pendugaan kurva regresi secara teoritik dan secara visual. Secara teoritik akan dibandingkan nilai koefisien determinasi ($R^2$) dan MSE pada setiap model yang didapatkan, sedangkan secara visual akan dibandingkan sebaran data bangkitan dengan kurva penduganya. Model yang bersifat *parsimoni* (sederhana) diharapkan sudah mampu menduga sebaran data dengan baik. Pada penelitian ini akan dilakukan iterasi dalam menentukan banyaknya titik knot dan orde yang digunakan untuk memperoleh kurva penduga dengan cara meminimumkan nilai *generalized cross validation* (GCV). Banyaknya titik knot yang dicobakan adalah satu, dua, dan tiga knot pada model regresi *spline truncated* orde 2 (linier), orde 3 (kuadratik), dan orde 4 (kubik) .

Adapun langkah-langkah pada penelitian ini sebagai berikut:

1. Membangkitkan data dengan satu variabel prediktor dan respon.

   Data variabel prediktor $(x)$ yang berdistribusi *Uniform* [0,3] dibangkitkan sebanyak 100 pengamatan dan dipetakan oleh fungsi:

1. f(x) = sin(4x)
2. f(x) = cos (4x).

Kemudian membangkitan $\varepsilon_i$ yang berdistribusi:

a. $\varepsilon \sim N(0{,}0.3)$

b. $\varepsilon \sim N(0{,}1)$

c. $\varepsilon \sim N(0{,}1.3)$

d. $\varepsilon \sim 0.95N(0{,}1) + 0.05N(5{,}1)$

e. $\varepsilon \sim 0.9N(0{,}1) + 0.1N(5{,}1)$.

Sehingga diperoleh data variabel respon $(y)$ melalui persamaan:

1. $y = \sin(4x) + \varepsilon_i$
2. $y = \cos(4x) + \varepsilon_i$.

3. Membuat diagram pencar antara variabel prediktor dengan variabel respon yang telah dibangkitkan untuk melihat pola sebaran data.
4. Membentuk model regresi *spline truncated* pada orde 2, orde 3, dan orde 4 untuk kemungkinan banyaknya knot adalah satu, dua, dan tiga knot berdasarkan nilai GCV.
5. Membandingkan kurva penduga pada beberapa kondisi galat yang berbeda.
6. Memilih model regresi *spline truncated* terbaik berdasarkan nilai MSE terkecil dan $R^2$ terbesar.
7. Membuat interpretasi dan kesimpulan.

# V. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya maka dapat diambil kesimpulan sebagai berikut:

1. Estimasi kurva *spline truncated* bergantung pada galat. Estimasi kurva *spline truncated* pada galat berdistribusi normal lebih baik dibandingkan pada galat yang mengandung pencilan baik pada fungsi sin (4x) maupun fungsi cos (4x).
2. Estimasi kurva *spline truncated* pada galat berdistribusi normal dengan varian yang kecil memberikan performa sangat baik bagi kurva estimasi untuk mendekati garis fungsi aslinya pada fungsi sin (4x) maupun fungsi cos (4x).

## DAFTAR PUSTAKA


Budiantara, I. N. 2006. Model Spline dengan Knots Optimal. *Jurnal Ilmu Dasar*. **7:** 77-85.

Budiantara, I.N. 2009. *Spline Dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Institut Teknologi Sepuluh November, Surabaya.

Chatterjee, S. 2006. *Regression Analysis by Example*. Jhon Wiley & Sons, New Jersey.

Drapper, N. & Smith, H. 1992. *Applied Regression Analysis*. Jhon Wiley & sons, New York.

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

Ghozali, I. 2011. *Aplikasi Analisis Multivariate Dengan Program SPSS.* Universitas Diponegoro, Semarang

Gujarati, D.N. 2003. *Basic Econometrics*. The McGraw-Hili Companies, New York.

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

Herawati, N. 2011. Regresi Spline untuk Pemodelan Bidang Kesehatan: Studi tentang Knot dan Selang Kepercayaan. *Jurnal ILMU DASAR*. **12**(2): 152-160.

Hidayat, R., dkk. 2017. Model Regresi Nonparametrik Dengan Pendekatan Spline Truncated, hlm. 203-352. Prosiding Seminar Nasional Institut Teknologi Sepuluh November, Surabaya.

Lee, T.C. 2002. On Algorithms for Ordinary Least Squares Regression Spline Fitting: A Comparative Study. *Journal Statist*. **72:** 647-663.

Wahba, G. 1990. *Spline Models For Observasion Data*. SIAM, Pensylvania.

Wan, M.P. 2000. A Comparison of Regression Spline Smoothing Procedures. *Computational Statistics*. **15:** 443-462.

Wu, H. & Zang, J.T. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*. Jhon Wiley & Sons, New Jersey.