



Estimator Spline Truncated of Semiparametric Regression *Estimator Regresi Semiparametrik Spline Truncated*

Narita Yuri Adrianingsih

Universitas Tribuana Kalabahi
Email: naritayuria98@gmail.com

Info Artikel

Sejarah Artikel:

Diterima: 27 Januari 2022

Direvisi: 31 Januari 2022

Dipublikasikan: Februari 2022

e-ISSN: 2089-5364

p-ISSN: 2622-8327

DOI: 10.5281/zenodo.6224937

Abstract:

Semiparametric regression is a combination of parametric and semiparametric regression. This regression is used on data whose pattern is known and partially the pattern is unknown. Ordinary Least Square (OLS), Maximum Likelihood Estimation (MLE), and Penalized Least Square (PLS) can be used to obtain Spline Truncated semiparametric regression curve estimators. This paper will explain about the use of Ordinary Least Square (OLS) to obtain a Spline Truncated Semiparametric Regression estimator.

Keywords: *OLS, Semiparametric Regression, Spline Truncated*

PENDAHULUAN

Pada analisis regresi terdapat 3 bentuk model regresi yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Ketiga model regresi ini mempunyai peranan yang penting dalam mengestimasi pola hubungan antara variabel respon dan variabel prediktor. Dalam regresi parametrik kurva regresi diasumsikan diketahui, dan diperlukan pengetahuan pada masa lalu tentang karakteristik data, sedangkan regresi nonparametrik kurva regresi tidak diketahui (Budiantara, 2005). Namun, dalam kasus-kasus tertentu variabel respon dapat memiliki hubungan yang diketahui dengan salah satu variabel prediktor, tetapi variabel prediktor lainnya tidak diketahui hubungannya, maka regresi semiparametrik dapat digunakan (Wahba, 1990).

Regresi nonparametrik mempunyai kelebihan, yaitu lebih fleksibel. Terdapat beberapa pendekatan yang ada dalam regresi nonparametrik, yaitu Kernel, Spline, Polinomial Lokal, deret Fourier, dan sebagainya (Budiantara, 2006). Salah satu pendekatan yang paling sering digunakan oleh peneliti-peneliti adalah pendekatan Spline Truncated, seperti yang telah dilakukan oleh Budiantara (2009), Dani *et al* (2020), Lestari & Budiantara (2020), Ramli *et al* (2020), Rosanti & Budiantara (2020), dan Setyowati *et al* (2020). Spline Truncated juga diaplikasikan secara campuran dengan pendekatan nonparametrik yang lainnya oleh beberapa peneliti, seperti yang telah dilakukan oleh Dewanti *et al* (2020), Octavanny *et al* (2020), dan Mariati *et al* (2020). Spline truncated mempunyai kelebihan yaitu mempunyai fleksibilitas yang tinggi,

sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik data. Spline truncated bersifat berubah-ubah pada sub-interval tertentu.

Regresi semiparametrik merupakan gabungan regresi parametrik dan regresi nonparametrik. Regresi semiparametrik ini pertama kali diperkenalkan oleh Wahba tahun 1990, yang meneliti tentang model hubungan semiparametrik dari penggunaan listrik di Amerika Serikat yang mengikuti pendekatan semiparametrik. Akhir-akhir ini, beberapa pengembangan penelitian tentang regresi semiparametrik adalah Budiantara (2020), Chamidah & Budiantara (2020), dan Khalil *et al* (2020).

Estimasi kurva regresi semiparametrik spline dapat diselesaikan dengan menggunakan *Ordinary Least Square* (OLS), *Maximum Likelihood Estimation* (MLE), *Penalized Least Square* (PLS). Dalam penelitian ini, diberikan penjelasan menggunakan metode OLS untuk mendapatkan estimator kurva regresi semiparametrik Spline *Truncated*.

METODE PENELITIAN

Tahapan estimasi regresi semiparametrik dengan pendekatan spline *truncated* univariabel adalah sebagai berikut:

(1) Diberikan data yang mengikuti model regresi semiparametrik sebagai berikut:

$$y_i = f(x_i) + h(z_i) + \varepsilon_i \quad (1)$$

(2) Menghampiri kurva regresi komponen parametrik dengan fungsi linier sebagai berikut:

$$f(x_i) = \beta_0 + \sum_{p=1}^P \beta_p x_{pi} \quad (2)$$

Apabila ditulis dalam bentuk matriks adalah sebagai berikut:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{12} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

notasi matriksnya adalah $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ (3)

dengan

$$\mathbf{y} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{12} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

(3) Menghampiri kurva regresi komponen nonparametrik menggunakan spline truncated dengan knot sebanyak r dan orde sebanyak M

$$h(z_i) = \sum_{j=1}^M a_j z_i^j + \sum_{k=1}^r b_k (z_i - \xi_k)_+^M \quad (4)$$

Apabila ditulis dalam bentuk matriks adalah sebagai berikut:

$$\begin{bmatrix} h(z_1) \\ h(z_2) \\ \vdots \\ h(z_n) \end{bmatrix} = \begin{bmatrix} z_1 & \cdots & z_1^M & (z_1 - \xi_1)_+^M & \cdots & (z_1 - \xi_r)_+^M & \vdots \\ z_2 & \cdots & z_2^M & (z_2 - \xi_1)_+^M & \cdots & (z_2 - \xi_r)_+^M & a_M \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & b_1 \\ z_n & \cdots & z_n^M & (z_n - \xi_1)_+^M & \cdots & (z_n - \xi_r)_+^M & \vdots \\ & & & & & & b_r \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_M \\ b_1 \\ \vdots \\ b_r \end{bmatrix}$$

notasi matriksnya adalah $\mathbf{y} = \mathbf{N}\mathbf{v}$ (5)

dengan

$$\mathbf{y} = [h(z_1), h(z_2), \dots, h(z_n)]^T$$

$$\mathbf{N} = \begin{bmatrix} z_1 & \cdots & z_1^M & (z_1 - \xi_1)_+^M & \cdots & (z_1 - \xi_r)_+^M \\ z_2 & \cdots & z_2^M & (z_2 - \xi_1)_+^M & \cdots & (z_2 - \xi_r)_+^M \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_n & \cdots & z_n^M & (z_n - \xi_1)_+^M & \cdots & (z_n - \xi_r)_+^M \end{bmatrix}$$

$$\mathbf{v} = [a_1, \dots, a_M, b_1, \dots, b_r]^T$$

(4) Menuliskan model regresi semiparametrik spline truncated dalam bentuk notasi matriks

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v} + \boldsymbol{\varepsilon} \quad (6)$$

(5) Menentukan error dari model

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{v} \quad (7)$$

(6) Mendapatkan estimasi untuk parameter dengan menggunakan optimasi OLS:

$$\min_{\boldsymbol{\beta}, \mathbf{v}} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v})\} \quad (8)$$

(7) Menyelesaikan optimasi dengan menggunakan derivative parsial.

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0, \frac{\partial l(\mathbf{v})}{\partial \mathbf{v}} = 0 \quad (9)$$

(8) Mendapatkan estimator regresi semiparametrik Spline Truncated

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{N}\hat{\mathbf{v}} \quad (10)$$

HASIL DAN PEMBAHASAN

Diberikan data berpasangan $(x_{1i}, x_{2i}, \dots, x_{pi}, z_i, y_i)$ yang memiliki hubungan, diasumsikan mengikuti model regresi semiparametrik. Variabel-variabel

$x_{1i}, x_{2i}, \dots, x_{pi}, z_i$ merupakan variabel prediktor, dan y_i merupakan variabel respon. Bentuk model regresi semiparametrik tersebut adalah:

$$\begin{aligned} y_i &= \mu(x_{1i}, x_{2i}, \dots, x_{pi}, z_i) + \varepsilon_i \\ &= \mu(x_i, z_i) + \varepsilon_i \end{aligned} \quad (11)$$

Dimana $x_i = x_{1i}, x_{2i}, \dots, x_{pi}$. Error random ε_i berdistribusi normal. Kurva regresi $\mu(x_i, z_i)$ diasumsikan bersifat

additive, sehingga dapat ditulis menjadi $\mu(x_i, z_i) = f(x_{1i}) + f(x_{2i}) + \dots + f(x_{pi}) + h(z_i)$

Kurva regresi $\mu(x_i, z_i)$ disebut dengan kurva regresi semiparametrik spline truncated. Persamaan dapat ditulis menjadi

$$\mu(x_i, z_i) = \sum_{p=1}^P f(x_{pi}) + h(z_i) \quad (12)$$

Komponen kurva regresi $\sum_{p=1}^P f(x_{pi})$

merupakan komponen kurva regresi linier berganda dan $h(z_i)$ merupakan komponen kurva regresi Spline Truncated.

Estimator untuk komponen kurva regresi linier berganda $\sum_{p=1}^P f(x_{pi})$ pada persamaan (12) apabila ditulis dalam bentuk notasi matriks akan menjadi sebagai berikut:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} \quad (13)$$

Estimator untuk komponen kurva regresi spline truncated $h(z_i)$ pada persamaan (12) apabila ditulis dalam bentuk notasi matriks akan menjadi sebagai berikut:

$$\mathbf{y} = \mathbf{N}\mathbf{v} \quad (14)$$

Dari persamaan (13) dan (14) maka persamaan (11) dapat disajikan dalam bentuk notasi matriks sebagai berikut:

$$\begin{aligned} y &= \sum_{p=1}^P f(x_{pi}) + h(z_i) + \varepsilon \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v} + \boldsymbol{\varepsilon} \end{aligned} \quad (15)$$

Dari persamaan (15) didapatkan persamaan sebagai berikut:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v} \quad (16)$$

Akan dilakukan penyelesaian untuk mendapatkan estimasi parameter pada persamaan (16) dengan menggunakan optimasi *Ordinary Least Square* (OLS).

$$\min_{\boldsymbol{\beta}, \mathbf{v}} \{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{v}) \}$$

Terlebih dahulu dilakukan penyederhanaan persamaan untuk mendapatkan estimasi parameter $\boldsymbol{\beta}$, yaitu sebagai berikut:

$$\begin{aligned} \mathbf{C} &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{v})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{v}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} - \mathbf{y}^T \mathbf{N}\mathbf{v} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \\ &\quad \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{N}\mathbf{v} - \mathbf{v}^T \mathbf{N}^T \mathbf{y} + \\ &\quad \mathbf{v}^T \mathbf{N}^T \mathbf{X}\boldsymbol{\beta} + \mathbf{v}^T \mathbf{N}^T \mathbf{N}\mathbf{v} \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} - 2\mathbf{v}^T \mathbf{N}^T \mathbf{y} \\ &\quad + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{N}\mathbf{v} \\ &\quad + \mathbf{v}^T \mathbf{N}^T \mathbf{N}\mathbf{v} \end{aligned} \quad (17)$$

Kemudian dilanjutkan penurunan pertama pada persamaan (17) dan disamadengankan nol.

$$\begin{aligned} \frac{\partial \mathbf{C}}{\partial \boldsymbol{\beta}} &= 0 \\ -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} + 2\mathbf{X}^T \mathbf{N}\mathbf{v} &= 0 \\ \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{N}\mathbf{v} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{N}\mathbf{v}) \end{aligned} \quad (18)$$

Kemudian dilanjutkan penyederhanaan persamaan untuk mendapatkan estimasi \mathbf{v} .

$$\begin{aligned} \mathbf{C} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{N}\mathbf{v})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{N}\mathbf{v}) \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} - 2\mathbf{v}^T \mathbf{N}^T \mathbf{y} \\ &\quad + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} + 2\mathbf{v}^T \mathbf{N}^T \mathbf{X}\boldsymbol{\beta} + \mathbf{v}^T \mathbf{N}^T \mathbf{N}\mathbf{v} \end{aligned} \quad (19)$$

Setelah itu dilanjutkan penurunan pertama persamaan (19) dan disamadengankan nol.

$$\begin{aligned} \frac{\partial \mathbf{C}}{\partial \mathbf{v}} &= 0 \\ -2\mathbf{N}^T \mathbf{y} + 2\mathbf{N}^T \mathbf{X}\hat{\boldsymbol{\beta}} + 2\mathbf{N}^T \mathbf{N}\hat{\mathbf{v}} &= 0 \\ \mathbf{N}^T \mathbf{N}\hat{\mathbf{v}} &= \mathbf{N}^T \mathbf{y} - \mathbf{N}^T \mathbf{X}\hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} &= (\mathbf{N}^T \mathbf{N})^{-1} (\mathbf{N}^T \mathbf{y} - \mathbf{N}^T \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned} \quad (20)$$

Dari hasil yang didapat pada persamaan (18) dan (20) terlihat bahwa masih mengandung parameter, maka untuk menyelesaikan persamaan tersebut digunakan metode substitusi. Untuk memudahkan perhitungan maka persamaan tersebut dapat ditulis sebagai berikut:

$$\hat{\beta} = D(y - N\hat{v}) \quad (21)$$

Dengan $D = (X^T X)^{-1} X^T$

$$\hat{v} = E(y - X\hat{\beta}) \quad (22)$$

Dengan $E = (N^T N)^{-1} N^T$

Langkah pertama substitusikan persamaan (22) ke persamaan (21) sehingga didapatkan:

$$\begin{aligned} \hat{\beta} &= D(y - N(Ey - EX\hat{\beta})) \\ &= Dy - DNEy + DNEX\hat{\beta} \\ \hat{\beta} - DNEX\hat{\beta} &= Dy - DNEy \end{aligned} \quad (23)$$

$$(I - DNEX)\hat{\beta} = Dy - DNEy$$

$$\hat{\beta} = (I - DNEX)^{-1}(Dy - DNEy)$$

$$\hat{\beta} = Oy$$

dimana $O = (I - DNEX)^{-1}(D - DNE)$

Kemudian substitusikan persamaan (23) ke persamaan (22) sehingga didapatkan sebagai berikut:

$$\begin{aligned} \hat{v} &= E(y - X\hat{\beta}) \\ &= Ey - EXOy \\ &= Py \end{aligned} \quad (24)$$

dimana $P = Ey - EXO$

Berdasarkan persamaan (23) dan (24), didapatkan estimasi fungsi semiparametric Spline Truncated sebagai berikut:

$$\begin{aligned} \hat{y} &= X\hat{\beta} + N\hat{v} \\ &= XOy + NPy \\ &= (XO + NP)y \end{aligned} \quad (25)$$

KESIMPULAN

Estimasi untuk model semiparemetrik Spline Truncated adalah sebagai berikut:

$$\begin{aligned} \hat{y} &= X\hat{\beta} + N\hat{v} \\ &= XOy + NPy \\ &= (XO + NP)y \end{aligned}$$

Dimana $O = (I - DNEX)^{-1}(D - DNE)$ dan

$$P = Ey - EXO$$

DAFTAR PUSTAKA

- Budiantara IN. 2005. *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik*. Berkala MIPA, 15(3). ITS Surabaya.
- Budiantara IN. 2009. *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa*

Kini dan Masa Mendatang. Pidato Pengukuhan untuk Jabatan Guru Besar dalam Bidang Ilmu Matematika Statistika dan Probabilitas, pada Jurusan Statistika, Fakultas MIPA. ITS, Surabaya.

- Budiantara IN. 2020. *Three form fourier series estimator semiparametric regression for longitudinal data*. Journal of Physics: Conference Series 1538 (1), 012058.
- Chamidah N, Budiantara IN. 2020. *Theoretical Study of Fourier Series Estimator in Semiparametric Regression for Longitudinal Data Based on Weighted Least Square Optimization*. 1st International Multidisciplinary Conference on Education, Technology and Engineering (IMCETE 2019)
- Dani A, Adrianingsih NY, Ainurrochmah. 2020. *Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi*. Jambura Journal of Probability and Statistics 1(2), 98-106.
- Dewanti P, Budiantara IN, Rumiati AT. 2020. *Modelling of SDG's Achievement in East Java Using Bi-Responses Nonparametric Regression with Mixed Estimator Spline Truncated and Kernel*. Journal of Physics: Conference Series 1562 (1), 012016.
- Khalil AA, Budiantara IN, Zain I. 2020. *Comparison of Linear and Quadratic Bi-response Semiparametric Regression Models Using Spline Truncated*. Journal of Physics: Conference Series 1511 (1), 012046.
- Lestari B & Budiantara IN. 2020. *Spline estimator and its asymptotic properties in miltiresponse nonparametric regression model*. Songklanakarin Journal of Science & Technology 42 (3).
- Mariati NPAM, Budiantara IN, Ratnasari V. 2020. *Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression*. Journal of Mathematics 2020.

- Octavanny MAD, Budiantara IN, Kuswanto H, Rahmawati DP. 2020. *Nonparametric Regression Model for Longitudinal Data with Mixed Truncated Spline and Fourier Series*. Abstract and Applied Analysis 2020.
- Ramli M, Ratnasari V, Budiantara IN. 2020. *Estimation of Matrix Variance-Covariance on Nonparametric Regression Spline Truncated for Longitudinal Data*. Journal of Physics: Conference Series 1562 (1), 012014.
- Rosanti IW & Budiantara IN. 2020. *Pemodelan Faktor-Faktor yang Mempengaruhi Morbiditas di Jawa Tengah Menggunakan Regresi Nonparametrik Spline Truncated*. Inferensi 3(2), 107-114.
- Setyowati DW, Rumiati AT, Budiantara IN. 2020. *Pemodelan Contraceptive Prevalence Rate (CPR) di Provinsi Sulawesi Selatan Menggunakan Regresi Nonparametrik Spline Truncated*. Jurnal Sains dan Seni ITS 9 (1), D72-D78.
- Wahba G. 1990. *Spline Models for Observational Data*. SIAM, Philadelphia