

3197-10567-1-SP Turnitin Naskah Awal

by 3197 10567

Submission date: 21-Nov-2022 05:37PM (UTC+0700)

Submission ID: 1960222121

File name: 3197-10567-1-SP_Turnitin_Naskah_Awal.docx (121.3K)

Word count: 1444

Character count: 8628

Estimasi Model Regresi Spline Kubik Tersegmen dengan Metode Penalized Least Square

No.artikel: 3197

Abstract

Nonparametric regression is used for data whose data pattern does not follow a parametric form. One of the estimators that can be developed is a segmented cubic spline which is able to show several segmentation changes in the data. This article examines the estimation of segmented cubic spline nonparametric regression models using the Penalized Least Square estimation criteria. The method involves knot points and smoothing parameters simultaneously. In addition, the model is used to analyze big data on BPJS claims based on patient age. The results show that the optimal model is at two-knot points, namely 26 and 52 with a smoothing parameter of 0.89. There are three segmentation changes from the cubic data, which consist of young people up to 26 years old, 26-52 years old, and 52 years and over.

Keywords: *Cubic Spline; Penalized Least Square; Segmentation.*

Abstrak

Regresi nonparametrik digunakan untuk data yang pola datanya tidak mengikuti bentuk parametrik. Salah satu estimator yang dapat dikembangkan adalah spline kubik tersegmen yang mampu menunjukkan beberapa segmentasi perubahan pada data. Artikel ini mengkaji estimasi model regresi nonparametrik spline kubik tersegmen melalui kriteria estimasi menggunakan Penalized Least Square. Metode tersebut melibatkan titik knot dan parameter penghalus secara bersamaan. Selain itu, model digunakan untuk menganalisis data besar klaim BPJS berdasarkan usia pasien. Hasil menunjukkan bahwa model optimal pada dua titik knot yaitu 26 dan 52 dengan parameter penghalus sebesar 0,89. Terdapat tiga segmentasi perubahan data secara kubik, yaitu usia muda hingga 26 tahun, usia 26-52 tahun, dan usia 52 tahun ke atas.

Kata Kunci: *Penalized Least Square; Segmentasi; Spline Kubik.*

1. Pendahuluan

Model regresi bertujuan untuk melihat pengaruh variabel prediktor terhadap variabel respon. Salah satu regresi yang telah banyak digunakan pada data riil adalah regresi parametrik. Asumsi pada regresi parametrik adalah bentuk pola data yang diasumsikan diketahui, misalnya linier, kuadratik, kubik, maupun bentuk lainnya. Akan tetapi, tidak semua data dapat diasumsikan mengikuti model regresi parametrik. Hal tersebut disebabkan karena terdapat data yang memiliki pola data yang tidak mengikuti pola parametrik atau ada segmentasi pola yang terjadi pada data. Oleh sebab itu, model regresi dikembangkan ke dalam regresi nonparametrik yang digunakan untuk data dengan bentuk fungsi regresinya belum diketahui¹. Beberapa estimator yang dikembangkan untuk mengestimasi fungsi regresi adalah estimator polinomial lokal², estimator kernel³, deret fourier⁴, dan estimator spline⁵.

Spline adalah suatu fungsi yang memuat titik knot pada variabel prediktor dan telah dikembangkan baik pada bentuk estimatornya maupun pada aplikasi. Beberapa penelitian yang terkait diantaranya regresi spline truncated pada data longitudinal⁶, spline linier penalized⁷, spline smoothing untuk data berkorelasi⁸, spline kuantil⁹, spline PCA¹⁰, dan beberapa bentuk spline lainnya. Pada aplikasi data riil, spline linier digunakan pada data diabetes¹¹, spline penalized linier pada data berat badan balita¹², spline kuantil

¹ Nyoman Budiantara et al., "Modeling Percentage of Poor People in Indonesia Using Kernel and Fourier Series Mixed Estimator in Nonparametric Regression."

² Chamidah et al., "Improving of Classification Accuracy of Cyst and Tumor Using Local Polynomial Estimator."

³ Budiantara et al., "Modeling Percentage of Poor People in Indonesia Using Kernel and Fourier Series Mixed in Nonparametric Regression."

⁴ Mardianto, Tjahjono, and Rifada, "Semiparametric Regression Based on Three Forms of Trigonometric Function in Fourier Series Estimator."

⁵ Islamiyati, "Spline Polynomial Truncated Dalam Regresi Nonparametrik."

⁶ Islamiyati, "Estimasi Kurva Regresi Spline Pada Data Longitudinal Dengan Metode Kuadrat Terkecil."

⁷ Islamiyati, Fatmawati, and Chamidah, "Penalized Spline Estimator with Multi Smoothing Parameters in Bi-Response Multi-Predictor Nonparametric Regression Model for Longitudinal Data."

⁸ Fernandes et al., "Spline Estimator for Bi-Responses and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data."

⁹ Aprilia, Islamiyati, and Anisa, "Estimasi Model Regresi Kuantil Spline Kuadratik Pada Data Trombosit Dan Hematokrit Pasien DBD."

¹⁰ Arifin, Islamiyati, and Raupong, "Kemampuan Estimator Spline Linear Dalam Analisis Komponen Utama."

¹¹ Islamiyati, Fatmawati, and Chamidah, "Changes in Blood Glucose 2 Hours after Meals in Type 2 Diabetes Patients Based on Length of Treatment at Hasanuddin University Hospital, Indonesia."

¹² Usrah and Islamiyati, "Analisis Perubahan Berat Badan Balita Dengan Estimator Penalized Spline Kuadratik."

kuadratik pada data DBD¹³, spline biprediktor pada data trombosit¹⁴ dan spline penalized robus pada data penjualan emas¹⁵.

Terdapat spline kubik yang merupakan salah satu estimator spline yang penting untuk dikaji karena bentuknya sebagai spline original. Peneliti telah mengembangkan spline smoothing pada orde kubik¹⁶, dan spline kubik pada konsep data cross section dan mengestimasi kurva regresi melalui pendekatan least square¹⁷. Pada perkembangan selanjutnya, spline telah dikembangkan pada sebuah kriteria *Penalized Least Square* (PLS) yang melibatkan titik knot pada fungsi *goodness of fit* dan parameter penghalus pada fungsi penalti. Keterlibatan titik knot yang mampu mengontrol ketepatan dan kehalusan kurva secara bersamaan menjadi keunggulan dari kriteria estimasi PLS. Oleh sebab itu, artikel ini menunjukkan bentuk segmentasi yang terbentuk pada spline kubik yang diestimasi dengan menggunakan PLS.

Kemampuan model regresi spline kubik yang tersegmentasi diaplikasikan pada data klaim BPJS berdasarkan usia. Klaim BPJS dimaksudkan besarnya klaim dari rumah sakit untuk pasien yang telah dirawat di rumah sakit. Dalam artikel, kecenderungan besar klaim ke BPJS berdasarkan usia ditunjukkan melalui segmentasi perubahan pola yang terbentuk dalam estimasi model regresi. Pada artikel ini, titik knot dan parameter penghalus bekerja secara simultan dalam kriteria estimasi yang dipilih berdasarkan nilai GCV minimum.

2. Metode

Data penelitian diperoleh dari catatan Rumah Sakit Pendidikan Universitas Hasanuddin Tahun 2020. Diperoleh data sebanyak 362 sampel dengan faktor usia sebagai prediktor dan faktor klaim BPJS sebagai respon. Penelitian ini merupakan kajian pengembangan model regresi nonparametrik spline tersegmentasi khusus pada estimasi model. Diketahui model regresi nonparametrik dinyatakan sebagai berikut:

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

Dengan y_i adalah faktor respon, $f(x_i)$ dinyatakan sebagai fungsi regresi yang belum diketahui bentuknya, hanya diasumsikan smooth, dan e_i adalah faktor error.

¹³ Puteri, Islamiyati, and Anisa, "Penggunaan Regresi Kuantil Multivariat Pada Perubahan Trombosit Pasien Demam Berdarah Dengue."

¹⁴ Islamiyati, "Regresi Spline Polynomial Truncated Biprediktor Untuk Identifikasi Perubahan Jumlah Trombosit Pasien Demam Berdarah Dengue."

¹⁵ Musafirah, Islamiyati, and Sunusi, "Estimation of Penalized Spline Linear Regression Models through Robust M Estimator."

¹⁶ Lestari et al., "Smoothing Parameter Selection Method for Multiresponse Nonparametric Regression Model Using Smoothing Spline and Kernel Estimators Approaches."

¹⁷ Ramdhani, Islamiyati, and Raupong, "Hubungan Faktor Kolesterol Terhadap Gula Darah Diabetes Dengan Spline Kubik Terbobot."

Selanjutnya, $f(x_i)$ dalam Persamaan (1) diestimasi melalui fungsi spline kubik sebagai berikut:

$$f(x_i) = \sum_{j=0}^3 a_j x_i^j + \sum_{k=1}^d a_{3+k} (x_i - K_k)_+^3 \quad (2)$$

Dengan x_i adalah faktor prediktor, a adalah koefisien regresi spline, K adalah titik knot dan $(x_i - K_k)_+^3$ adalah fungsi truncated yang menunjukkan segmentasi pada data dan dapat dinyatakan sebagai berikut:

$$(x_i - K_k)_+^3 = \begin{cases} (x_i - K_k)^3 & ; x_i \geq K_k \\ 0 & ; x_i < K_k \end{cases} \quad (3)$$

3. Hasil dan Diskusi

Model regresi spline kubik tersegmen adalah regresi spline yang memuat fungsi truncated berdasarkan titik knot. Pada penelitian ini, data yang dianalisis adalah data pertumbuhan tinggi badan berdasarkan usia balita. Oleh sebab itu, berdasarkan Persamaan (1) dan (2), model regresi spline kubik tersegmen yang bersesuaian dengan data dapat dinyatakan sebagai berikut:

$$y_i = \sum_{j=0}^3 a_j x_i^j + \sum_{k=1}^d a_{3+k} (x_i - K_k)_+^3 + e_i, i = 1, 2, \dots, 250 \quad (4)$$

Selanjutnya Persamaan (4) dapat dinyatakan ke dalam bentuk matriks sebagai berikut:

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \boldsymbol{\varepsilon}$$

Estimasi terhadap parameter \mathbf{a} dilakukan melalui kriteria estimasi PLS yang dinyatakan sebagai berikut:

$$PLS = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + l \boldsymbol{\alpha}^T \mathbf{D}\mathbf{a}$$

dengan $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\mathbf{a}$, sehingga

$$PLS = (\mathbf{y} - \mathbf{X}\mathbf{a})^T (\mathbf{y} - \mathbf{X}\mathbf{a}) + l \boldsymbol{\alpha}^T \mathbf{D}\mathbf{a}$$

dan diperoleh:

$$PLS = \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{X} \mathbf{a} + l \boldsymbol{\alpha}^T \mathbf{D}\mathbf{a} \quad (5)$$

Selanjutnya, Persamaan (5) diturunkan terhadap parameter \mathbf{a} seperti berikut ini:

$$\frac{\partial PLS}{\partial \mathbf{a}} = -2\mathbf{X}^T \mathbf{y} + 2\boldsymbol{\alpha}^T \mathbf{X}^T \mathbf{X} + 2l \boldsymbol{\alpha}^T \mathbf{D}$$

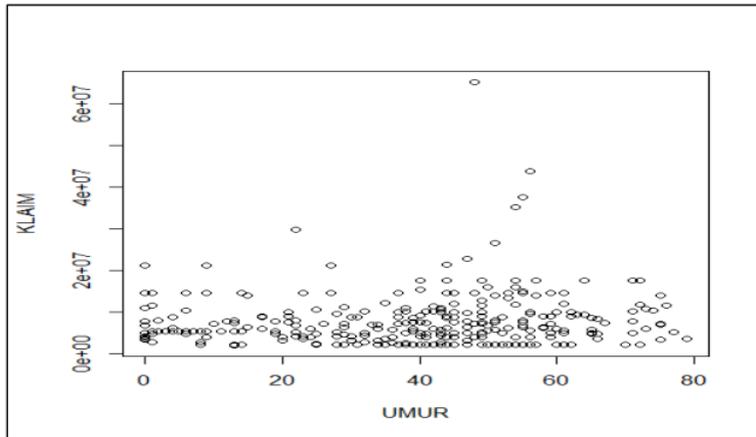
Berdasarkan hasil penguraian, diperoleh estimasi terhadap parameter regresi spline kubik tersegmen adalah $\hat{\mathbf{a}}$ seperti berikut:

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X} + l \mathbf{D})^{-1} \mathbf{X}^T \mathbf{y} \quad (6)$$

Berdasarkan hasil Persamaan (5) dan model pada Persamaan (4), estimasi model regresi spline kubik tersegmen dengan PLS dapat dituliskan sebagai berikut:

$$\hat{y} = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4(x - K_1)_+^3 + K + a_{3+d}(x - K_d)_+^3 \quad (7)$$

Hasil estimasi parameter dalam Persamaan (6) dan taksiran model pada Persamaan (7) digunakan untuk menganalisis data klaim BPJS di Rumah Sakit Pendidikan Universitas Hasanuddin. Untuk plot data awal, data diplotkan dalam scatter plot untuk menunjukkan pola hubungan antara peubah seperti pada Gambar 1.



Gambar 1. Scatter plot antara faktor klaim BPJS dengan umur pasien

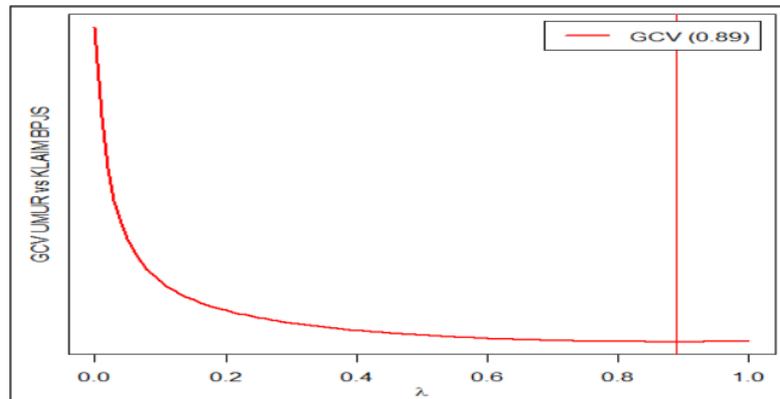
Plot yang dihasilkan pada Gambar 1 menunjukkan bahwa pola data tidak mengikuti bentuk parametrik. Hal tersebut ditunjukkan oleh titik-titik sampel yang tidak beraturan sehingga penggunaan regresi parametrik kurang tepat untuk data. Dalam artikel ini, model yang digunakan adalah regresi nonparametrik dengan estimator spline kubik seperti pada Persamaan (7).

Selanjutnya, jumlah titik knot yang dicobakan pada data adalah 1, 2, dan 3 knot. Pemilihan titik knot dan parameter penghalus optimal berdasarkan nilai GCV minimum ditunjukkan pada Tabel 1.

Tabel 1. Nilai GCV minimum pada titik knot dan parameter penghalus optimal

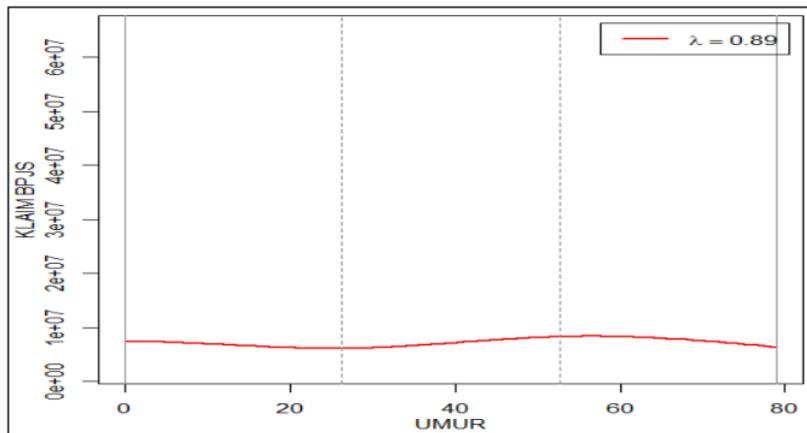
K_1	λ	K_1	K_2	λ	K_1	K_2	K_3	λ	GCV
39	1,5								4,11+E13
		26	52	0,89					4,09+E13
					20	40	60	0,59	4,12+E13

Berdasarkan Tabel 1, diperoleh nilai GCV pada penggunaan 1, 2 dan 3 titik knot. Titik knot optimal berada pada 2 titik knot yaitu titik knot K_1 sebesar 26 dan K_2 sebesar 52. Hal ini disebabkan karena nilai GCV yang diperoleh untuk 2 titik knot tersebut adalah nilai GCV minimum yaitu sebesar $4,09+E13$. Titik knot optimal yang diperoleh bersesuaian dengan nilai parameter penghalus sebesar $\lambda = 0,89$ seperti yang ditunjukkan pada Gambar 2.



Gambar 2. Nilai GCV minimum pada beberapa nilai parameter penghalus

Nilai parameter penghalus yang ditunjukkan pada Gambar 2 dimulai dari 0 hingga 1 dan diperoleh nilai GCV minimum terletak pada parameter penghalus $\lambda = 0,89$. Hal ini menunjukkan bahwa parameter penghalus yang digunakan dalam memodelkan data klaim BPJS adalah 0,89.



Gambar 3. Estimasi kurva regresi spline kubik tersegmen pada data klaim BPJS

1 Berdasarkan hasil estimasi model regresi nonparametrik **spline** kubik tersegmentasi pada data dengan 2 titik knot, yaitu $K_1 = 26$, $K_2 = 52$, dan $\lambda = 0,89$ seperti berikut ini:

$$\hat{y} = 4.406.358 + 8.911.581x + 4.519.716x^2 + 9.657.163x^3 + 6.633.433(x - 26)_+^3 + 1.922.660(x - 52)_+^3$$

Berdasarkan model optimal yang diperoleh menunjukkan bahwa terdapat tiga segmentasi pola kubik yang terjadi pada data klaim BPJS berdasarkan umur pasien. Pola pertama adalah pada usia pasien muda mencapai 26 tahun, terlihat klaim BPJS rumah sakit mengalami penurunan secara kubik. Pola kedua pada pasien adalah usia 26-52 tahun yang terlihat ada segmentasi pola lain terbentuk. Terdapat klaim BPJS mengalami kenaikan secara kubik yang mengindikasikan banyaknya pasien peserta BPJS yang berada dalam rentang usia tersebut. Selanjutnya, pola terakhir pada usia 52 tahun ke atas menunjukkan suatu segmentasi penurunan klaim BPJS yang mengindikasikan bahwa pasien peserta BPJS di Rumah Sakit Pendidikan Universitas Hasanuddin tidak sebanyak pada kelompok usia 26-52 tahun.

4. Kesimpulan

Model regresi nonparametrik dengan spline kubik yang tersegmentasi melibatkan titik-titik knot dan pada saat parameter regresi diestimasi dalam kriteria PLS, maka titik knot tersebut bekerja simultan dengan parameter penghalus dalam mengestimasi model. Taksiran model regresi pada data klaim BPJS menunjukkan bahwa terdapat tiga segmentasi perubahan besar klaim BPJS oleh rumah sakit. Segmentasi pertama pada pasien muda hingga 26 tahun dengan besar klaim BPJS yang turun perlahan secara kubik, segmentasi kedua pada pasien usia 26-52 tahun dengan besar klaim BPJS mengalami peningkatan dan segmentasi ketiga pada pasien usia 52 tahun ke atas dengan besar klaim BPJS kembali menurun.

3197-10567-1-SP Turnitin Naskah Awal

ORIGINALITY REPORT

7%

SIMILARITY INDEX

7%

INTERNET SOURCES

2%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1 journal.unhas.ac.id 4%
Internet Source

2 repository.its.ac.id 3%
Internet Source

3 stiepena.ac.id 1%
Internet Source

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On