

5.3197-10640-1-CE Turnitin Naskah Akhir

by 3197 10640

Submission date: 01-Dec-2022 09:37AM (UTC+0700)

Submission ID: 1967828638

File name: 5.3197-10640-1-CE_Turnitin_Naskah_Akhir.docx (80.6K)

Word count: 1428

Character count: 8751

Estimasi Model Regresi Spline Kubik Tersegmen dengan Metode Penalized Least Square

Estimation of Segmented Cubic Spline Regression Models by Penalized Least Square Method

No.artikel 3197

Abstract

Nonparametric regression is used for data whose data pattern is non-parametric. 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 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 bentuk non 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 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.

Pendahuluan

Model regresi bertujuan untuk melihat pengaruh variabel prediktor terhadap variabel respon. Salah satu model regresi yang telah banyak digunakan pada data riil adalah regresi parametrik. Bentuk pola data pada regresi parametrik diasumsikan diketahui, misalnya linier, kuadratik, kubik, maupun bentuk lainnya. Akan tetapi, tidak semua data dapat diasumsikan mengikuti model regresi parametrik. 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 kernel*³, deret *fourier*⁴, dan *estimator spline*⁵.

Estimator 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 penalized*⁷, *spline*

¹ I. Nyoman Budiantara et al., "Modeling Percentage of Poor People in Indonesia Using Kernel and Fourier Series Mixed Estimator in Nonparametric Regression," *Investigación Operacional* 40, no. 4 (September 5, 2019): 538-50, <http://www.invoperacional.uh.cu/index.php/InvOp/article/view/700>.

² Nur Chamidah et al., "Improving of Classification Accuracy of Cyst and Tumor Using Local Polynomial Estimator," *Telkomnika (Telecommunication Computing Electronics and Control)* 17, no. 3 (2019): 1492-1500, <https://doi.org/10.12928/TELKOMNIKA.V17I3.12240>.

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

⁴ M. F. F. Mardianto, E. Tjahjono, and M. Rifada, "Semiparametric Regression Based on Three Forms of Trigonometric Function in Fourier Series Estimator," *Journal of Physics: Conference Series* 1277, no. 1 (July 2019): 012052, <https://doi.org/10.1088/1742-6596/1277/1/012052>.

⁵ Anna Islamiyati, "Spline Polynomial Truncated Dalam Regresi Nonparametrik," *Jurnal Matematika, Statistika Dan Komputasi* 14, no. 1 (February 11, 2018): 54-60, <https://doi.org/10.20956/jmsk.v14i1.3538>.

⁶ Anna Islamiyati, "Taksiran Kurva Regresi Spline Pada Data Longitudinal Dengan Kuadrat Terkecil," *Jurnal Matematika, Statistika Dan Komputasi* 11, no. 1 (2014): 77-83, <https://doi.org/10.20956/jmsk.v11i1.3432>.

⁷ Anna Islamiyati, Fatmawati, and Nur Chamidah, "Penalized Spline Estimator with Multi Smoothing Parameters Bi-Response Multi-Predictor Nonparametric Regression Model for Longitudinal Data," *Songklanakar Journal of Science and Technology* 42, no. 4 (2020): 897-909, <http://www.scopus.com/inward/record.url?scp=85085869891&partnerID=8YFLogxK>.

smoothing untuk data berkorelasi ⁸, *spline kuantil* ⁹, *spline PCA* ¹⁰, dan beberapa bentuk *spline* lainnya. Pada aplikasi data riil, *spline* linier digunakan pada data diabetes ¹¹, *spline penalized* linier pada data berat badan balita ¹², *spline kuantil kuadrat* pada data DBD ¹³, *spline biprediktor* pada data trombosit¹⁴ dan *spline penalized robus* pada data penjualan emas ¹⁵.

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 kubik ¹⁶, dan *spline* kubik pada konsep data cross section dan mengestimasi kurva regresi melalui pendekatan least square ¹⁷. 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 ¹⁸.

⁸ Adji Achmad Rinaldo Fernandes et al., "Spline Estimator for Bi-Responses and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data," *Journal of Mathematics and Statistics* 11, no. 2 (October 8, 2015): 61-69, <https://doi.org/10.3844/jmssp.2015.61.69>.

⁹ Bunga Aprilia et al., "Estimasi Model Regresi Kuantil Spline Kuadrat Pada Data Trombosit Dan Hematokrit Pasien DBD," *ESTIMASI: Journal of Statistics and Its Application* 1, no. 2 (2020): 58-64, <https://doi.org/10.20956/ejsa.v1i2.9264>.

¹⁰ Samsul Arifin, Anna Islamiyati, and Raupong Raupong, "Kemampuan Estimator Spline Linear Dalam Analisis Komponen Utama," *ESTIMASI: Journal of Statistics and Its Application* 1, no. 1 (2020): 40-47, <https://doi.org/10.20956/ejsa.v1i1.9262>.

¹¹ Anna Islamiyati, Fatmawati, and Nur Chamidah, "Changes in Blood Glucose 2 Hours After Meals in Type 2 Diabetes Patients Based on Length of Treatment at Hasanuddin University Hospital, Indonesia," *Rawal Medical Journal* 45, no. 1 (January 1, 2020): 31-34, <http://www.scopus.com/inward/record.url?scp=85081662228&partnerID=8YFLogxK>.

¹² Muhammad Jayzul Usrah, Anna Islamiyati, and Anisa, "Analisis Perubahan Berat Badan Balita Dengan Estimator Penalized Spline Kuadrat," *ESTIMASI: Journal of Statistics and Its Application* 3, no. 2 (July 2, 2022): 70-75, <https://doi.org/10.20956/ejsa.vi.11459>.

¹³ Widya Nauli Amalia Puteri, Anna Islamiyati, and Anisa Anisa, "Penggunaan Regresi Kuantil Multivariat Pada Perubahan Trombosit Pasien Demam Berdarah Dengue," *ESTIMASI: Journal of Statistics and Its Application*, 2020, 1-9, <https://doi.org/10.20956/ejsa.v1i1.9224>.

¹⁴ Anna Islamiyati, "Regresi Spline Polynomial Truncated Biprediktor Untuk Identifikasi Perubahan Jumlah Trombosit Pasien Demam Berdarah Dengue," *Al-Khwarizmi: Jurnal Pendidikan Matematika Dan Ilmu Pengetahuan Alam* 7, no. 2 (2019): 97-112, <https://doi.org/10.24256/jpmipa.v7i2.799>.

¹⁵ Anna Islamiyati Musafirah, "Estimation of Penalized Spline Linear Regression Models through Robust M Estimator," *International Journal of Engineering and Information Systems (IJEAIS)* 5, no. 11 (November 28, 2021): 166-68, <http://ijeais.org/wp-content/uploads/2021/11/abs/IJAAR211120.html>.

¹⁶ B. Lestari et al., "Smoothing Parameter Selection Method for Multiresponse Nonparametric Regression Model Using Smoothing Spline and Kernel Estimators Approaches," *Journal of Physics: Conference Series* 1397, no. 1 (December 2019): 012064, <https://doi.org/10.1088/1742-6596/1397/1/012064>.

¹⁷ Zhazha Alifkhamulki Ramdhani, Anna Islamiyati, and Raupong Raupong, "Hubungan Faktor Kolesterol Terhadap Gula Darah Diabetes Dengan Spline Kubik Terbobot," *ESTIMASI: Journal of Statistics and Its Application* 1, no. 1 (2020): 32-39, <https://doi.org/10.20956/ejsa.v1i1.9252>.

¹⁸ Islamiyati, Fatmawati, and Chamidah, "Penalized Spline Estimator with Multi Smoothing Parameters in Bi-Response Multi-Predictor Nonparametric Regression Model for Longitudinal Data."

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 *Generalized Cross Validation* (GCV) minimum.

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. Diketahui model regresi nonparametrik dinyatakan sebagai berikut:

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

Dengan y_i adalah faktor respon, $f(x_i)$ dinyatakan sebagai fungsi regresi yang belum diketahui bentuknya, hanya diasumsikan *smooth*, dan ε_i adalah faktor *error*. Selanjutnya, $f(x_i)$ dalam Persamaan (1) diestimasi melalui fungsi *spline* kubik sebagai berikut:

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

Dengan x_i adalah faktor *prediktor*, α 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:

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

Hasil dan Diskusi

Model regresi *spline* kubik tersegmen adalah regresi *spline* yang memuat fungsi truncated berdasarkan titik knot. Berdasarkan Persamaan (1) dan (2), model regresi *spline* kubik tersegmen yang bersesuaian dengan data dapat dinyatakan sebagai berikut:

$$y_i = \sum_{j=0}^3 \alpha_j x_i^j + \sum_{k=1}^d \alpha_{3+k} (x_i - K_k)_+^3 + \varepsilon_i, i = 1, 2, \dots, 362 \quad (4)$$

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

$$y = X\alpha + \varepsilon$$

Estimasi terhadap paramater α dilakukan melalui kriteria estimasi PLS yang dinyatakan sebagai berikut:

$$PLS = \varepsilon^T \varepsilon + \lambda \alpha^T D \alpha$$

dengan $\varepsilon = y - X\alpha$, sehingga

$$PLS = (y - X\alpha)^T (y - X\alpha) + \lambda \alpha^T D \alpha$$

dan diperoleh:

$$PLS = y^T y - 2\alpha^T X^T y + \alpha^T X^T X \alpha + \lambda \alpha^T D \alpha \quad (5)$$

Selanjutnya, Persamaan (5) diturunkan terhadap paramater α seperti berikut ini:

$$\frac{\partial PLS}{\partial \alpha} = -2X^T y + 2X^T X \alpha + 2\lambda D \alpha$$

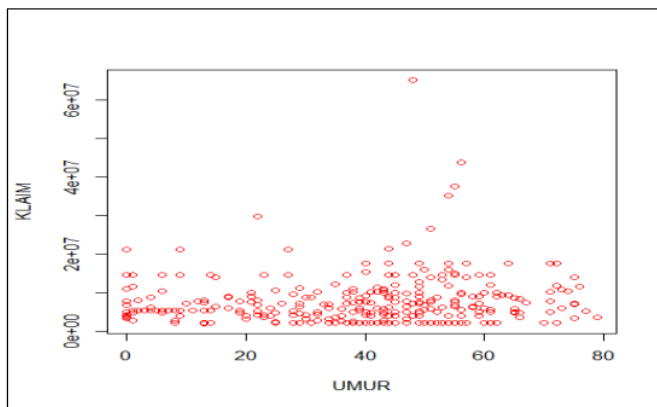
Dan jika hasil turunan α sama dengan nol, maka diperoleh estimasi terhadap parameter regresi *spline* kubik tersegmen adalah $\hat{\alpha}$ seperti berikut:

$$\hat{\alpha} = (X^T X + \lambda D)^{-1} X^T y \quad (6)$$

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

$$\hat{y} = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 (x - K_1)_+^3 + \dots + \alpha_{3+d} (x - K_d)_+^3 \quad (7)$$

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

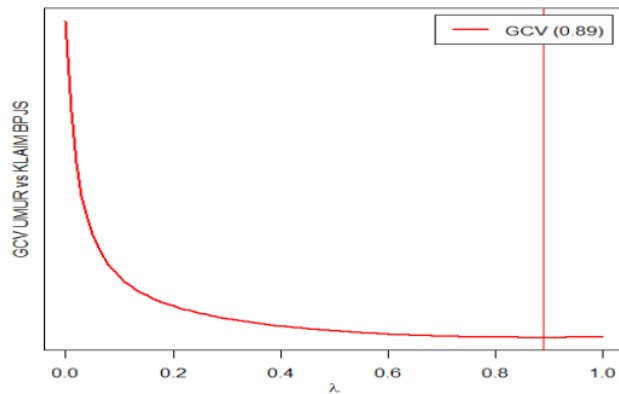
Plot yang dihasilkan pada Gambar 1 menunjukkan bahwa pola data tidak mengikuti bentuk parametrik. Hal tersebut ditunjukkan oleh titik-titik sampel yang tidak beraturan sehingga penggunaan regresi parametrik kurang tepat untuk data. Dalam artikel ini, model yang digunakan adalah regresi nonparametrik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

Selanjutnya, jumlah titik knot yang dicobakan pada data adalah 1, 2, dan 2 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

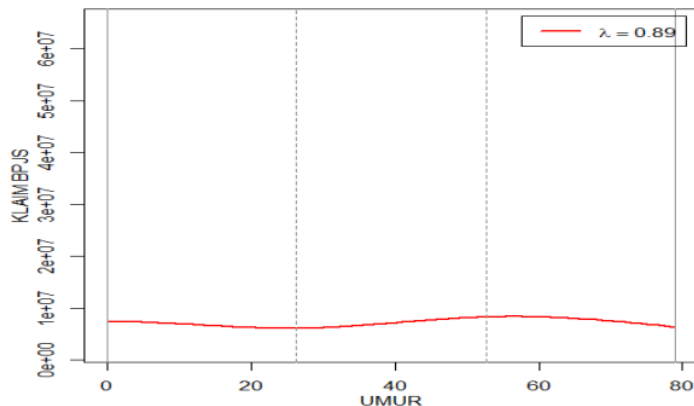
Berdasarkan Tabel 1, diperoleh nilai GCV pada penggunaan 1, 2 dan 3 titik knot. Titik knot optimal berada pada 2 titik knot yaitu titik knot K_1 sebesar 26 dan K_2 sebesar 52. Hal ini disebabkan karena nilai GCV yang diperoleh untuk 2 titik knot tersebut adalah nilai GCV minimum yaitu sebesar 4,09+E13. 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. Selanjutnya hasil estimasi model regresi nonparametrik spline kubik tersegmen pada data dengan 2 titik knot, yaitu $K_1 = 26$, $K_2 = 52$, dan $\lambda = 0,89$ seperti berikut ini:

$$\hat{y} = 4.406.358 + 8.911.581x + 4.519.716x^2 + 9.657.163x^2 + 6.633.433(x - 26)_+^3 + 1.922.660(x - 52)_+^3 \quad (8)$$



Gambar 3. Estimasi Kurva Regresi *Spline* Kubik Tersegmen pada Data Klaim BPJS

Persamaan (8) merupakan model yang bersesuaian dengan Gambar 3 yang 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.

Kesimpulan

Model regresi nonparametrik dengan spline kubik yang tersegmen melibatkan titik-titik knot dan pada saat parameter regresi diestimasi dalam kriteria PLS, maka titik knot tersebut bekerja simultan dengan parameter penghalus dalam melakukan estimasi 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.

5.3197-10640-1-CE Turnitin Naskah Akhir

ORIGINALITY REPORT

6%

SIMILARITY INDEX

6%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

ejournal.iainpalopo.ac.id

Internet Source

4%

2

journal.unhas.ac.id

Internet Source

2%

Exclude quotes On

Exclude bibliography On

Exclude matches < 2%