

Received: 10 May 2023
Revised: 25 December 2023
Accepted: 30 December 2023
Published: 31 December 2023

Regresi Nonparametrik *Spline Truncated* untuk Memodelkan Tingkat Pengangguran Terbuka di Pulau Kalimantan

Rewani Ega Tanaty, Darnah*, Sifriyani

Program Studi Statistika, Universitas Mulawarman Samarinda

*Email korespondensi: darnah.98@gmail.com

Abstract

Nonparametric regression is a statistical technique employed to discern the relationship pattern between a predictor variable and a response variable in the absence of prior information about the form of the regression function or when the pattern of the regression curve is unknown. Truncated spline nonparametric regression represents an approach for aligning data, considering the curve's smoothness. It possesses continuous segmented characteristics, offering flexibility and adeptly accommodating the explanation of local data function features. The study aims to identify the influencing factors on the open unemployment rate in Kalimantan Island during 2020. Additionally, it seeks to derive a spline truncated nonparametric regression model for Open Unemployment Rate data in Kalimantan Island for the same year. The study employs a nonparametric regression model with a spline truncated method, determining optimal knot points based on the minimum Generalized Cross Validation (GCV) value. The study reveals that the most effective spline truncated nonparametric regression model features two knot points. Significant factors influencing the open unemployment rate include the labor force participation rate, school year expectations, regional gross domestic product at current prices, and the human development index.

Keywords: GCV, Nonparametric Regression, Spline Truncated, Open Unemployment Rate, Knot Point.

Abstrak

Regresi nonparametrik adalah metode statistika yang digunakan untuk mengidentifikasi pola hubungan antara variabel prediktor dan variabel respons tanpa memiliki informasi sebelumnya tentang bentuk fungsi regresi atau kurva regresinya. Metode ini memiliki karakteristik tersegmen kontinu, memberikan fleksibilitas dan efektivitas dalam menyesuaikan diri untuk menjelaskan ciri lokal dari fungsi data. Tujuan penelitian ini adalah

untuk menyelidiki faktor-faktor yang memengaruhi tingkat pengangguran terbuka di Pulau Kalimantan pada tahun 2020. Selain itu, penelitian ini bertujuan untuk mengembangkan model regresi nonparametrik spline truncated untuk data tingkat Pengangguran Terbuka di Pulau Kalimantan pada tahun tersebut. Hasil penelitian menunjukkan bahwa model regresi nonparametrik spline truncated terbaik dapat dicapai dengan menggunakan dua titik knot yang optimal berdasarkan nilai GCV minimum. Faktor-faktor yang secara signifikan memengaruhi tingkat pengangguran terbuka adalah tingkat partisipasi angkatan kerja, harapan lama sekolah, produk domestik regional bruto atas dasar harga berlaku, dan indeks pembangunan manusia.

Kata kunci: GCV, Regresi Nonparametrik, Spline Truncated, Tingkat Pengangguran Terbuka, Titik Knot

PENDAHULUAN

Analisis regresi merupakan metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respon dan variabel prediktor (Draper & Smith, 1992; Gujarati, 2006; William, dkk., 2023). Salah satu tujuan dari analisis regresi adalah untuk menentukan bentuk hubungan antara variabel respon dan variabel prediktor. Pada umumnya terdapat tiga pendekatan dalam mengestimasi kurva regresi, yaitu pendekatan parametrik, nonparametrik, dan semiparametrik (Nurhuda, dkk., 2022). Pendekatan dalam regresi nonparametrik telah banyak dikembangkan antara lain menggunakan *spline*, kernel, deret fourier (Eubank, 1998; Darnah, dkk., 2019). Salah satu model regresi dengan pendekatan nonparametrik yang sangat sering digunakan untuk melakukan estimasi terhadap kurva regresi adalah regresi *spline* (Pratama, dkk., 2022). Model regresi nonparametrik dengan pendekatan *spline* dapat diaplikasikan dalam berbagai bidang salah satunya pada tingkat pengangguran terbuka karena datanya tidak mengikuti pola tertentu. Tingkat pengangguran terbuka merupakan perbandingan antara banyaknya pencari kerja dengan angkatan kerja. Pengangguran terbuka tercipta karena adanya penambahan tenaga kerja yang banyak sedangkan lapangan pekerjaan yang tersedia sedikit. Pengangguran terbuka merupakan seseorang yang berusia 15 tahun atau lebih yang tidak memiliki pekerjaan dan sedang mencari pekerjaan, dan ada juga yang sudah mempunyai pekerjaan, tetapi belum mulai bekerja. Tingginya angka pengangguran dipengaruhi oleh Pendapatan Domestik Regional Bruto (PDRB) dan 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 (Sukirno, 1994). Penelitian ini menggunakan regresi nonparametrik *spline truncated* untuk memodelkan tingkat pengangguran terbuka di Kalimantan pada tahun 2020.

Masalah pengangguran selalu menjadi permasalahan krusial yang dihadapi oleh semua negara. Pulau Kalimantan merupakan salah satu pulau yang memiliki kepadatan penduduk yang cukup tinggi di Indonesia. Data Tingkat Pengangguran Terbuka (TPT) pada Badan Pusat Statistik (BPS) di Pulau Kalimantan mengalami penurunan setiap tahunnya sedangkan pada tahun 2020 mengalami peningkatan sebesar 0,41%. Pada tahun 2019 tingkat pengangguran terbuka di Pulau Kalimantan sebesar 4,75%, dan pada tahun 2020 sebesar 5,15% (Badan Pusat Statistika, 2020). Oleh karena itu, peneliti ingin mengetahui faktor-faktor yang memengaruhi tingkat pengangguran terbuka di Pulau Kalimantan tahun 2020 menggunakan model regresi nonparametrik *spline truncated* sehingga dapat menjadi salah satu referensi pengambilan kebijakan untuk mengendalikan TPT di Pulau Kalimantan.

METODOLOGI

Bahan dan Data

Penelitian ini mengambil lokasi tiap kabupaten dan kota yang berada di Pulau Kalimantan, yaitu sebanyak 56 kabupaten/kota. Data diperoleh dari *website* Badan Pusat Statistik Tahun 2020 dan variabel penelitian dapat dilihat pada TABEL 1.

TABEL 1. Variabel Penelitian

Notasi	Variabel	Skala
y	Tingkat Pengangguran Terbuka	Rasio
x_1	Tingkat Partisipasi Angkatan Kerja	Rasio
x_2	Harapan Lama Sekolah	Rasio
x_3	Produk Domestik Regional Bruto Atas Dasar Harga Berlaku	Rasio
x_4	Indeks Pembangunan Manusia	Rasio

Regresi Nonparametrik dengan Estimasi Spline Truncated

Regresi *spline truncated* adalah suatu pendekatan dengan tetap memperhitungkan kemulusan kurva, *spline* merupakan model polinomial yang tersegmen. Sifat tersegmen inilah sehingga memberikan fleksibilitas yang lebih baik daripada model polinomial biasa. Sifat ini juga memungkinkan model regresi *spline* dapat menyesuaikan diri secara efektif terhadap karakteristik lokal data. Pendekatan *spline truncated* mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun dengan bantuan titik-titik knot, serta kurva yang dihasilkan relatif *smooth* (Eubank, 1999). Fungsi *spline truncated* dengan satu variabel prediktor adalah

$$f(x_i) = \sum_{l=0}^m \beta_l x_i^l + \sum_{h=1}^r \beta_{m+h} (x_i - k_h)_+^m \tag{2}$$

dengan:

$$(x_i - k_h)_+^m = \begin{cases} (x_i - k_h)^m, & x_i \geq k_h \\ 0, & x_i < k_h \end{cases} \tag{3}$$

Nilai m pada persamaan (3) merupakan derajat polinomial. Kurva polinomial derajat satu disebut dengan kurva linier, derajat dua disebut dengan kurva kuadratik, dan derajat tiga disebut dengan kurva kubik.

Jika terdapat satu variabel respon dengan lebih dari satu variabel prediktor maka model regresi nonparametrik *spline truncated* dapat dituliskan sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^p \sum_{l=1}^m \beta_{jl} x_{ji}^l + \sum_{j=1}^p \sum_{h=1}^r \beta_{j,m+h} (x_{ji} - k_{hj})_+^m + \varepsilon_i \tag{4}$$

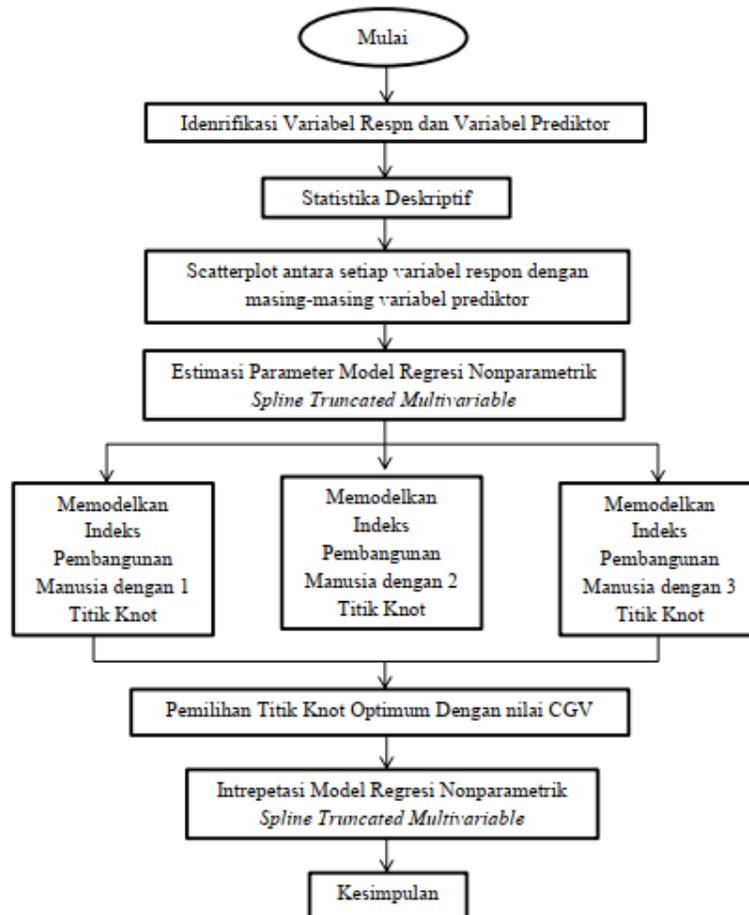
Pemilihan Titik Knot Optimal

Pemilihan titik knot optimal dilakukan untuk mendapatkan model *spline truncated* terbaik. Nilai GCV dapat ditentukan dengan rumus berikut (Wahba, 1990):

$$GCV = \frac{MSE(K)}{[n^{-1}trace(I - A(K))^2]} \tag{5}$$

Metode Penelitian

Metode yang digunakan pada penelitian ini yaitu regresi nonparametrik dengan pendekatan *spline truncated* untuk satu titik knot, dua titik knot, serta tiga titik knot. Pemilihan titik knot optimal menggunakan metode *Generalized Cross Validation* (GCV). Diagram alir untuk penelitian ini ditunjukkan pada GAMBAR 1.



GAMBAR 1. Diagram Alir Analisis Data

HASIL DAN PEMBAHASAN

Statistik Deskriptif

Sebelum melakukan pemodelan terhadap Tingkat Pengangguran Terbuka dengan pendekatan regresi nonparametrik *spline truncated*, dilakukan analisis ststistika deskriptif terlebih dahulu. Statistika deskriptif bertujuan untuk memberikan gambaran tentang variabel penelitian (Walpole,

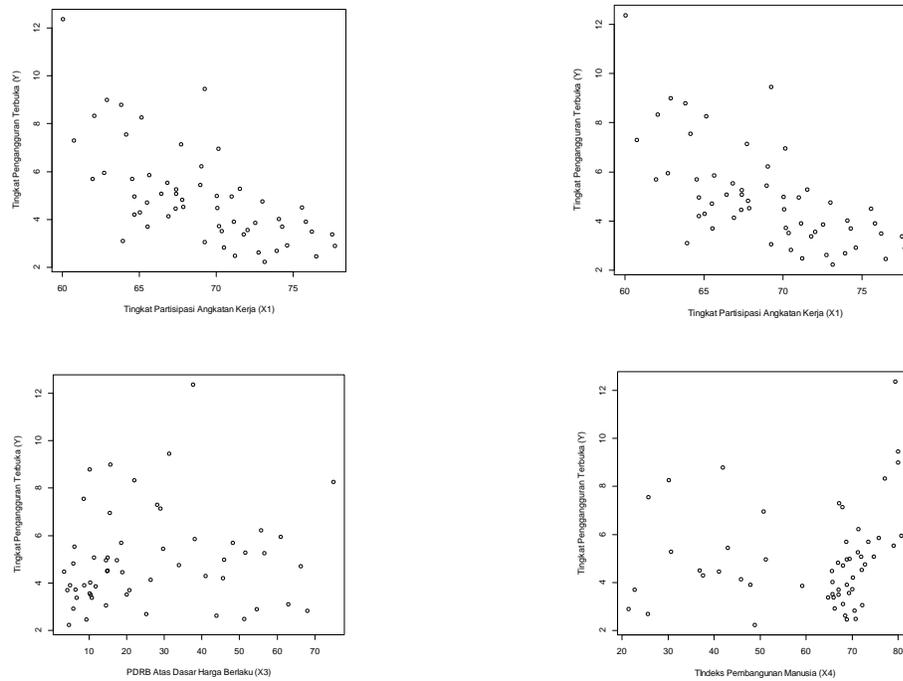
1995; Darnah, 2013). Hasil analisis statistik deskriptif variabel penelitian dapat dilihat pada TABEL 2.

TABEL 2. Statistik Deskriptif

Variabel	Jumlah Data	Rata-rata	Minimum	Maksimum	Standar Deviasi
y	56	4,957	2,240	12,360	2,032
x_1	56	68,999	60,050	77,730	4,496
x_2	56	12,644	11,150	14,940	0,807
x_3	56	26,362	3,300	75,040	20,018
x_4	56	61,122	21,390	80,770	16,479

Scatterplot

Sebelum melakukan pemodelan pada data, perlu dilakukan visualisasi terhadap pola data. Hasil *scatterplot* berfungsi untuk mengetahui apakah terbentuk suatu pola tertentu atau tidak. Jika pada *scatterplot* terbentuk suatu pola tertentu seperti linier, kuadratik, kubik atau pola lainnya, maka dalam pemodelan digunakan pendekatan regresi parametrik. Namun, apabila *scatterplot* tidak membentuk pola tertentu, maka digunakan pendekatan regresi nonparametrik. Jika dalam *scatterplot* terdapat sebagian membentuk pola tertentu dan sebagian tidak membentuk pola apapun, maka digunakan pendekatan regresi semi-parametrik.



GAMBAR 2. *Scatterplot* antara variabel prediktor dengan variabel respon

Berdasarkan GAMBAR 2, pola hubungan antara jumlah kejadian Tingkat Pengangguran Terbuka dengan variabel Tingkat Partisipasi Angkatan Kerja, Harapan lama sekolah, PDRB Atas Dasar Harga Berlaku, dan Indeks Pembangunan Manusia di Pulau Kalimantan menunjukkan pola data yang tidak membentuk kurva tertentu. Pada *scatterplot* jumlah kejadian Tingkat Pengangguran Terbuka dengan keempat variabel prediktor Tingkat Partisipasi Angkatan Kerja, Harapan lama sekolah, Produk Domestik Regional Bruto Atas Dasar Harga Berlaku, dan Indeks Pembangunan

Manusia terdapat beberapa data pencilan sehingga regresi nonparametrik tepat digunakan dalam mengestimasi jumlah kejadian Tingkat Pengangguran Terbuka.

Model Regresi Nonparametrik *Spline Truncated*

Model terbaik regresi nonparametrik spline truncated diperoleh berdasarkan nilai GCV minimum. Perbandingan nilai GCV dengan 1, 2, dan 3 titik knot dapat dilihat pada TABEL 3.

TABEL 3. Perbandingan Nilai GCV 1, 2, dan 3 Titik Knot

Titik Knot	Nilai GCV Minimum
1 Titik Knot	15,081
2 Titik Knot	13,160
3 Titik Knot	14,442

TABEL 3 menunjukkan bahwa nilai GCV minimum adalah 13,160, yaitu pada model dengan dua titik knot sehingga titik knot optimal yang dapat digunakan sebagai model terbaik adalah regresi nonparametrik *spline truncated* dengan dua titik knot, yaitu:

$$\hat{y} = -1,859x_1 + 1,031(x_1 - 61,132) - 4,170(x_1 - 76,286) - 2,228x_2 + 9,154(x_2 - 11,382) - 2,164(x_2 - 14,630) - 3,547x_3 + 6,943(x_3 - 7,692) + 4,083(x_3 - 69,183) + 2,041x_4 + 4,845(x_4 - 25,025) - 1,020(x_4 - 75,922) \quad (1)$$

Uji Signifikansi Parameter

Setelah mendapatkan model terbaik, maka dilakukan pengujian signifikansi parameter regresi. Terdapat dua uji signifikansi parameter yang dilakukan yakni uji signifikansi parameter secara simultan dan uji secara parsial. Berikut bentuk hipotesis uji simultan:

$$\begin{aligned} H_0 & : \beta_{1,1} = \beta_{1,2} = \beta_{1,3} = \beta_{1,4} = \dots = \beta_{4,1} = \beta_{4,2} = \beta_{4,3} = \beta_{4,4} = 0 \\ H_1 & : \text{Minimal terdapat satu } \beta_{j(1+h)} \neq 0; \text{ dimana } j = 1, 2, \dots, 4 \text{ dan } h = 1, 2, 3 \end{aligned}$$

Hasil analisis menunjukkan nilai F_{hitung} statistik uji sebesar 14,45977 dengan p -value sebesar $2,421785 \times 10^{-13}$. Bila p -value dibandingkan dengan tingkat signifikansi yang digunakan yakni 0,05 maka dapat diambil keputusan bahwa H_0 ditolak sehingga disimpulkan bahwa minimal terdapat satu variabel yang memberikan pengaruh signifikan terhadap model. Selanjutnya, dilakukan uji parsial untuk mengetahui variabel mana saja yang memberikan pengaruh signifikan terhadap model. Hasil uji parsial dapat dilihat pada TABEL 4.

TABEL 4. Pengujian Parameter Model Regresi Secara Parsial

	Variabel Parameter	Estimasi Parameter	<i>p</i> – value	<i>t</i> – hitung	Keputusan
<i>y</i>	β_0	6,044	0,208	1,275	Tidak Signifikan
	$\beta_{1,1}$	-1,859	0,003	-3,138	
<i>x</i> ₁	$\beta_{1,2}$	1,031	0,005	2,919	Signifikan
	$\beta_{1,3}$	-4,170	0,168	1,401	
	$\beta_{2,1}$	-2,228	0,186	1,342	
<i>x</i> ₂	$\beta_{2,2}$	9,154	0,267	-1,122	Signifikan
	$\beta_{2,3}$	-2,164	0,000	-5,207	
	$\beta_{3,1}$	-3,547	0,683	-0,410	
<i>x</i> ₃	$\beta_{3,2}$	-3,547	0,683	-0,410	Signifikan
	$\beta_{3,3}$	6,943	0,000	-12,839	
	$\beta_{4,1}$	4,083	0,266	1,124	
<i>x</i> ₄	$\beta_{4,2}$	2,041	0,006	2,859	Signifikan
	$\beta_{4,3}$	4,845	0,005	-2,927	

TABEL 4 menunjukkan bahwa variabel yang signifikan memengaruhi Tingkat Pengangguran Terbuka di Pulau Kalimantan adalah tingkat partisipasi angkatan kerja, harapan lama sekolah, PDRB atas dasar harga berlaku, dan indeks pembangunan manusia. Model regresi nonparametrik *spline truncated* dengan 2 titik knot beserta variabel berpengaruh dapat dilihat pada persamaan (2):

$$\hat{y} = -1,859x_1 + 1,031(x_1 - 61,132)_+^1 - 4,170(x_1 - 76,286)_+^1 - 2,228x_2 + 9,154(x_2 - 11,382)_+^1 - 2,164(x_2 - 14,630)_+^1 - 3,547x_3 + 6,943(x_3 - 7,692)_+^1 + 4,083(x_3 - 69,183)_+^1 + 2,041x_4 + 4,845(x_4 - 25,025)_+^1 - 1,020(x_4 - 75,922)_+^1 + \varepsilon_i \tag{2}$$

Nilai koefisien determinasi (*R*²) sebesar 80,14 persen yang berarti bahwa tingkat partisipasi angkatan kerja, harapan lama sekolah, PDRB atas dasar harga berlaku, dan indeks pembangunan manusia sudah cukup menjelaskan 80,14 persen variasi Tingkat Pengangguran Terbuka di Pulau Kalimantan dan 19,86 persen variasi lainnya dijelaskan oleh variabel lain yang tidak digunakan dalam model regresi.

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Rambe, dkk. (2019) yang menyatakan bahwa PDRB berpengaruh terhadap Tingkat Pengangguran Terbuka. Selain itu, hasil penelitian Suhendra & Wicaksono (2016) juga menunjukkan bahwa harapan lama sekolah atau pendidikan berpengaruh negatif terhadap Tingkat Pengangguran Terbuka.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan dalam penelitian ini dapat disimpulkan bahwa model regresi nonparametrik *spline truncated* terbaik pada Tingkat Pengangguran Terbuka di Pulau Kalimantan Tahun 2020 adalah 2 titik knot dengan nilai GCV minimum adalah 13,160 dan *R*² sebesar 80,14%. Faktor-faktor yang memengaruhi Tingkat Pengangguran Terbuka di Pulau Kalimantan Tahun 2020 berdasarkan model regresi nonparametrik *spline truncated* terbaik adalah Tingkat Partisipasi Angkatan Kerja, Harapan Lama Sekolah, Produk Domestik Regional Bruto Atas Dasar Harga Belaku, dan Indeks Pembangunan Manusia. Penelitian selanjutnya dapat menggunakan metode *Cross Validation* (CV) atau *Unbiased Risk* (UBR) dalam pemilihan titik knot optimal.

REFERENSI

- Badan Pusat Statistik. (2020). *Profil Tingkat Pengangguran Terbuka*. Jakarta Pusat : Badan Pusat Statistik.
- Darnah, Chamidah, & Utoyo. (2019). Modeling of Material Mortality and Indant Mortality Cases in East Kalimantan using Poisson Regression Approach Based on Local Linear Estimator. *IOP Conference Series: Earth and Environmental Science*, 243, 1-7. doi:10.1088/1755-1315/243/1/012023.
- Darnah, A. N. (2013). *Biostatistika 1*. Jakarta: Halaman Moeka.
- Draper, N. R. & Smith, H. 1992. *Applied Regression Analysis*, Second Edition. John Wiley and sons, Inc. New York.
- Eubank, R. (1998). *Spline Smoothing and Nonparametric Regression*. Mareel Dekker: New York.
- Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing 2nd*. New York: Marcel Dekker Inc.
- Gujarati, D. N., 2006. *Dasar-dasar Ekonometrika Jilid 1 dan 2*. Edisi Ketiga. Erlangga. Jakarta.
- Karimuse, W. Y., Darnah, & Siringoringo, M. (2023). Pendekatan Regresi Nonparametrik Kernel pada Data IHSG Periode Januari 2020 – Desember 2021. *Statistika*: 23(1), 1-7.
- Nurhuda, G., Wasono, & Darnah. (2022). Pemodelan Regresi Nonparametrik Berdasarkan Estimator Spline Truncated pada Data Simulasi. *Jurnal Mataematika, Statistika, dan Komputasi*, 19(1): 172-182. DOI: 10.20956/j.v19i1.21534
- Pratama, M. H., Sifriyani, & Darnah. (2022). Regresi Nonparametrik Multivariabel dengan Pendekatan Spline Truncated Pada Kasus Tuberculosis. *Statistika*, 22(1), 87-93.
- Rambe, R. C., Prihanto, P. H., & Hardiani. 2019. Analisis Faktor-Faktor yang Mempengaruhi Pengangguran Terbuka di Provinsi Jambi. *e-Jurnal Ekonomi Sumberdaya dan Lingkungan*, 8(1): 54-67
- Serena, N., Sifriyani, & Darnah. (2021). Aplikasi Pendekatan Spline Truncated Dalam Model Gwr Pada Pencemaran Derah Aliran Sungai Mahakam. *Jurnal Bina Wakya*, 16(2):6439-6446.
- Suhendra, I. & Wicaksono, B.H. (2016). Tingkat Pendidikan, Upah, Inflasi, dan Pertumbuhan Ekonomi Terhadap Pengangguran di Indonesia. *Jurnal Ekonomi-Qu*, 6(1), 1-17.
- Sukirno. (1994). *Pengantar Ekonomi Makro*. PT. Raja Grasindo Persada. Jakarta.
- Wahba, G. (1990). *Spline Models for Observation Data*. Society for Industrial and Applied Mathematics. Philadelphia: Pennsylvania.
- Walpole, R. E. (1995). *Pengantar Statistika*, edisi ke-3. Jakarta: Gramedia Pustaka Utama.