

PEMODELAN IHK DENGAN REGRESI NONPARAMETRIK SPLINE MULTIVARIABEL

Azizatun Nisa¹, Makkulau¹, Lilis Laome¹, Edi Cahyono¹, Norma Muhtar¹

¹Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Halu Oleo, Kendari, Indonesia
E-mail: aziza.statistik016@gmail.com

ABSTRAK

Indeks Harga Konsumen (IHK) merupakan salah satu indikator ekonomi yang dapat memberikan informasi mengenai perkembangan harga barang dan jasa yang dibayar oleh konsumen atau masyarakat, khususnya masyarakat perkotaan sehingga secara umum mempengaruhi rata-rata harga yang tercipta antara produsen dengan konsumen. Karena pola data yang berubah-ubah berdasarkan selang waktu tertentu tidak tepat memakai regresi linear dan menggunakan lebih dari satu variabel prediktor oleh karena itu dibutuhkan suatu penelitian untuk mencari model terbaik sehingga mampu menjelaskan fenomena terkait menggunakan pendekatan yang sesuai dengan pola hubungan data. Salah satu metode yang dapat digunakan adalah Regresi Nonparametrik Spline. Spline merupakan potongan polinomial tersegmen yang memiliki titik perpaduan bersama yang menunjukkan terjadinya perubahan-perubahan perilaku kurva. Tujuan dari penelitian ini adalah untuk mengetahui pemodelan regresi spline multivariabel pada data Indeks Harga Konsumen. Berdasarkan model terbaik dari IHK di Indonesia adalah pada tiga titik knot dengan nilai GCV minimum 14,07 serta R-square (R^2) adalah 0.73 yang berarti kemampuan variabel respon dalam menjelaskan variabel terikatnya sebesar 73% sedangkan sisanya dipengaruhi oleh variabel lain diluar penelitian.

Kata Kunci: Indeks Harga Konsumen, spline multivariabel, GCV

Pendahuluan

Analisis regresi merupakan sebuah metode statistika yang memberikan penjelasan tentang pola hubungan (model) antara dua variabel atau lebih (Draper & Smith, 1992). Variabel-variabel tersebut dikelompokkan menjadi variabel respon (dependen) dan variabel prediktor (independen). Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus dalam arti termuat dalam suatu ruang fungsi tertentu, sehingga mempunyai sifat fleksibilitas yang tinggi (Winarti & Sony, 2010).

Dalam analisis regresi nonparametrik spline, jika terdapat satu variabel respon dan satu variabel prediktor maka regresi tersebut dinamakan regresi nonparametrik spline univariabel. Sebaliknya, apabila terdapat satu variabel respon dengan lebih dari satu variabel prediktor maka regresi tersebut disebut regresi nonparametrik spline multivariabel (Budiantara, 2005). Regresi nonparametrik spline multivariabel lebih baik dibandingkan dengan spline univariabel untuk beberapa variabel prediktor. Hal ini dapat menghasilkan fungsi

regresi yang lebih fleksibel dengan yang digunakan.

Estimasi parameter model spline bertujuan untuk mendapatkan model regresi nonparametrik Spline yang akan digunakan dalam analisis. Pada penelitian ini, metode yang digunakan untuk mengestimasi parameter model regresi nonparametrik Spline adalah metode kuadrat terkecil atau sering disebut dengan metode *Ordinary Least Square* (OLS). Dalam penelitian ini akan diteliti mengenai faktor yang diduga mempengaruhi IHK menggunakan metode regresi dengan pendekatan nonparametrik spline dan menggunakan *Generalized Cross Validation* (GCV) sebagai metode pemilihan titik knot optimal.

2.1 Regresi Nonparametrik Spline Multivariabel

Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Spline merupakan potongan polinomial yang memiliki sifat tersegmen sehingga, model spline memiliki fleksibilitas yang tinggi dan memiliki kemampuan sangat baik untuk menangani data yang perilakunya

berubah pada sub interval tertentu (Budiantara, 2009). Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval-interval yang berbeda disebut knot. Adapun model dari regresi spline adalah:

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_i + \sum_{k=1}^m \beta_{j+k} (x_i - k_k)^q + \varepsilon_i \quad (1)$$

dimana:

- y_i : variabel respon
- β_0 : Konstanta
- β_j : parameter model
- x_i : variabel predictor
- β_{j+k} : koefisien regresi untuk variabel X dan pada knot ke-k
- k_k : titik knot ke-k
- ε_i : *error* acak yang diasumsikan identik, independen, dan $N(0, \sigma^2)$

Dalam analisis regresi nonparametrik spline, jika terdapat satu variabel respon dan satu variabel prediktor maka regresi tersebut dinamakan regresi nonparametrik spline univariabel. Sebaliknya, apabila terdapat satu variabel respon dengan lebih dari satu variabel prediktor maka regresi tersebut disebut regresi nonparametrik spline multivariabel. Adapun model dari spline multivariabel adalah:

dimana:

$$\begin{aligned} y &= f(x_1) + f(x_2) + \dots + f(x_p) + \varepsilon \\ f(x_1) &= \beta_{01} + \beta_{11}x_1 + \beta_{21}(x_1 - k_{11})^1 \\ f(x_2) &= \beta_{02} + \beta_{12}x_2 + \beta_{22}(x_2 - k_{12})^1 \\ &\vdots \\ f(x_p) &= \beta_{0p} + \beta_{1p}x_p + \beta_{2p}(x_p - k_{1p})^1 \end{aligned}$$

dengan

$$f_1(x) = \begin{cases} \beta_0 + \beta_1 x & ; x < k \\ \beta_0 + \beta_1 x + \beta_2 (x - k) & ; x \geq k \end{cases}$$

keterangan:

- y : Variabel respon
- $f(x_i)$: fungsi regresi spline
- ε : *error* acak yang diasumsikan identik, independen, dan $N(0, \sigma^2)$
- $\beta_{01}, \beta_{02}, \dots, \beta_{0p}$: konstanta
- $k_{11}, k_{12}, \dots, k_{1p}$: titik knot
- x_p : variabel predictor
- $\beta_{11}, \beta_{12}, \dots, \beta_{1p}$: parameter variabel
- k : banyak knot

Metode yang umumnya digunakan untuk mengestimasi parameter regresi nonparametrik pada spline adalah OLS. Prinsip dari metode ini adalah dengan meminimumkan galat (*error*) yang dihasilkan oleh model sehingga diharapkan model regresi menjelaskan data dengan baik. Mengingat galat yang dihasilkan model regresi pada setiap amatan dapat bernilai negatif dan positif, maka untuk menghindari penjumlahan yang bernilai negatif, dicari jumlah dari kuadrat nilai galat (Putri, 2018). Adapun estimasi parameternya adalah:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

2.2 Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pada data. Model regresi spline terbaik tergantung pada titik knot optimal (Eubank, 1998). Salah satu metode untuk mencari titik knot optimal adalah GCV. Titik knot optimal diperoleh dari nilai GCV yang minimum. Kriteria lain yang dapat digunakan sebagai ukuran kinerja atas penaksir yang baik adalah GCV yaitu:

$$GCV(k) = \frac{MSE(k)}{(n^{-1} \text{trace}[I - A(k)])^2} \quad (3)$$

keterangan:

- I : matriks identitas
- n : jumlah pengamatan
- (2) $A(k)$: matriks $X(X'X)^{-1}X'$
- $MSE(k)$: $n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

2.3 Uji Signifikansi Parameter

Pengujian parameter ini dilakukan untuk menentukan variabel prediktor mana saja yang memiliki hubungan nyata dengan variabel respon. Untuk regresi nonparametrik spline terdapat dua tahap pengujian yaitu uji serentak dan uji individu.

- a. Uji Serentak adalah uji signifikansi parameter model secara keseluruhan atau untuk mengetahui apakah semua variabel prediktor yang dimasukkan ke dalam model memberikan pengaruh secara bersama-sama. Statistik uji yang digunakan adalah statistik uji F.
- b. Uji Individu Setelah dilakukan uji serentak, maka perlu diketahui secara individu parameter mana yang signifikan dan mana yang tidak signifikan. Statistik uji yang digunakan adalah statistik uji t.

Pengujian asumsi residual pada model yang dianalisis dengan menggunakan Regresi Nonparametri Spline, harus memenuhi asumsi residual identik, independen, dan berdistribusi normal.

- a. Uji Identik. Pengujian asumsi residual identik bertujuan mendeteksi adanya kasus heteroskedastisitas adalah upaya untuk mengurangi kerugian bagi efisiensi estimator. Cara mendeteksi dengan menggunakan Uji *Bruesch Pagan*.
- b. Uji Independen. Pengujian independensi residual bertujuan untuk mendeteksi apakah terdapat korelasi antar residual atau yang biasa disebut dengan autokorelasi. Cara mendeteksi dengan menggunakan uji *Durbin Watson*.
- c. Uji Distribusi Normal. Pengujian asumsi distribusi normal dilakukan untuk mendeteksi apakah residual berdistribusi normal. Pengujian ini dilakukan dengan menggunakan uji *Liliefors*.

2.4 Kriteria Keباikan Model

Salah satu ukuran yang sering digunakan untuk mengetahui kebaikan suatu model yaitu koefisien determinasi atau R^2 . Koefisien ini menunjukkan seberapa besar persentase variasi variabel prediktor yang digunakan dalam model mampu menjelaskan variasi variabel respon. Model yang baik adalah model yang dapat menjelaskan variabilitas dari variabel respon dengan baik (R^2 tinggi). Nilai R^2 dapat dihitung dengan rumus sebagai berikut:

$$R^2 = \frac{SSR}{SST} \times 100\% \quad (4)$$

keterangan:

SSR : *Sum Square Regression*

SST : *Sum Square Total*

Besaran nilai R^2 tidak pernah negatif. (Gujarati, 2003). Menghitung nilai koefisien determinasi merupakan salah satu rangkaian penting dalam menganalisis model data berpasangan melalui regresi. Apabila dengan hanya satu peubah bebas (X) diperoleh nilai koefisien determinasi yang cukup besar maka cukup kita berbicara model RLS saja. Tetapi apabila nilai koefisien determinasinya kecil maka dianjurkan untuk menggunakan metode regresi berganda atau melakukan transformasi data. (Soleh, 2005).

3. Hasil dan pembahasan

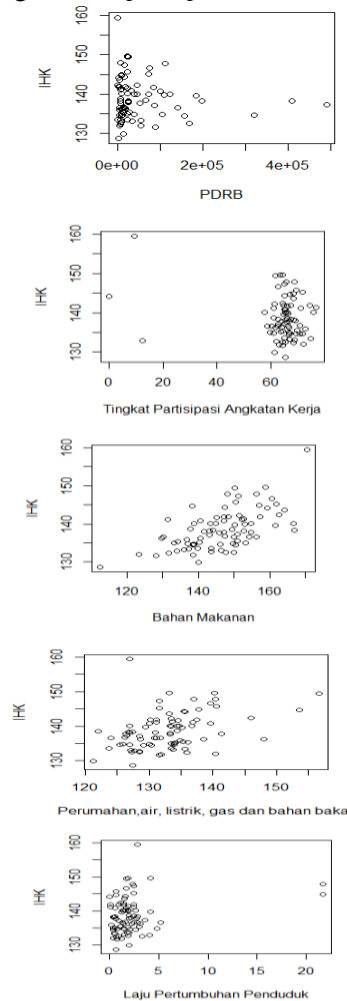
Berikut adalah statistika deskriptif data IHK yang bertujuan menampilkan karakteristik

dari data berdasarkan ukuran-ukuran nilai statistika. Seperti pada tabel 3.1 berikut.

Tabel 3.1 Data IHK

Variabel	Rata-rata	Minimum	Maksimum
Y	138,71	128,73	159,44
X_1	53.310,00	46,78	490.766,47
X_2	64,01	0,08	76,92
X_3	146,30	112,50	170,40
X_4	133,00	121,29	156,66
X_5	2,46	0,04	21,70

Mengetahui hubungan antara IHK dengan faktor-faktor yang diduga mempengaruhi merupakan langkah awal dalam melakukan pemodelan dengan menggunakan pendekatan regresi nonparametrik spline. Berikut merupakan identifikasi hubungan antara IHK dengan faktor-faktor yang diduga mempengaruhi. Seperti pada Gambar 3.1.



Gambar 3.1 Scatterplot IHK (y) dengan variabel-variabel prediktor

Berdasarkan Gambar 3.1 menunjukkan bahwa *Scatterplot* antara IHK dengan variabel-variabel yang diduga berpengaruh tidak menunjukkan adanya kecenderungan membentuk pola tertentu. Sehingga metode regresi nonparametrik spline sangat digunakan untuk pemodelan karena keseluruhan plot tidak membentuk pola tertentu, maka regresi nonparametrik spline dapat digunakan dalam pemodelan. Model regresi nonparametrik spline untuk 1 titik knot ditunjukkan pada Persamaan berikut:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (X_1 - k_1)^+ + \hat{\beta}_3 X_2 + \hat{\beta}_4 (X_2 - k_2)^+ + \hat{\beta}_5 X_3 + \hat{\beta}_6 (X_3 - k_3)^+ + \hat{\beta}_7 X_4 + \hat{\beta}_8 (X_4 - k_4)^+ + \hat{\beta}_9 X_5 + \hat{\beta}_{10} (X_5 - k_5)^+ + \varepsilon$$

keterangan:

\hat{y} : y duga

k : titik knot

X : variabel prediktor

β : parameter variabel

ε : error

Begitu pula untuk spline dengan menggunakan 2 titik knot dan 3 titik knot. Titik knot optimum untuk satu knot, dua knot, dan tiga knot didapatkan, maka langkah selanjutnya adalah membandingkan nilai GCV dari masing masing nilai knot optimum yang telah didapatkan. Berikut merupakan perbandingan GCV berbagai titik knot yang ditunjukkan oleh Tabel 3.2 berikut

Tabel 3.2 Perbandingan nilai GCV

Knot	GCV
Satu knot	14,074
Dua Knot	14,665
Tiga Knot	14,07

Dari nilai GCV yang dihasilkan dengan berbagai titik knot, dapat dilihat bahwa nilai GCV yang paling minimum, adalah dihasilkan oleh pemilihan tiga titik knot sehingga tiga titik knot merupakan pemodelan terbaik. Selanjutnya akan di uji signifikansi parameter pada model terbaik.

Hasil pengujian signifikansi parameter secara serentak untuk tiga titik knot disajikan pada Tabel 3.3 berikut.

Tabel 3.3 Hasil Signifikansi Parameter Uji Serentak Tiga Titik Knot

Sumber Variansi	Derajat Kebebasan (<i>df</i>)	Jumlah Kuadrat (JK)	Kuadrat Tengah (KT)	F_{hitung}	$Pr(>F)$
	Regresi	20	1794,40	89,72	5,92
Galat	61	924,77	15,16		
Total	81	2719,17			

Dari table diatas diketahui bahwa p-value adalah sebesar 0,00 karena nilai ini kurang dari nilai 0,05 Sehingga dapat disimpulkan bahwa terdapat satu parameter yang berpengaruh signifikan terhadap variabel respon. Sedangkan padapengujian parameter uji induvidu dengan pengujian signifikansi dilakukan dengan cara perbandingan nilai p-value dengan nilai α . terdapat 13 dari 21 parameter yang tidak signifikan karena memiliki nilai p-value yang lebih besar dari 0,05. Meskipun terdapat variabel yang tidak signifikan terhadap IHK, namun pemodelan kembali dengan menghapus variabel tidak perlu dilakukan karena variabel yang tidak berpengaruh signifikan terhadap IHK masih tetap dapat dimasukkan kedalam model hanya tidak perlu diinterpretasikan karena pengaruhnya tidak besar terhadap IHK.

Pengujian asumsi residual dilakukan untuk mengetahui kelayakan suatu model regresi. Analisis regresi nonparametrik spline memiliki asumsi identik, independen dan berdistribusi normal (IIDN). Apabila model regresi tidak memenuhi asumsi tersebut maka model tersebut tidak layak digunakan untuk memodelkan variabel respon. Berikut adalah Tabel 3.4

Tabel 3.4 Tabel Asumsi Residual

Model	Asumsi Residual			R^2	MSE
	Identik	Independen	Berdistribusi Normal		
Tiga titik knot	✓	✓	✓	73%	7,786

Berdasarkan Tabel 3.4 dapat kita lihat bahwa model dari tiga titik knot yaitu Tingkat partisipasi angkatan kerja (X_2), Bahan makanan (X_3), laju pertumbuhan penduduk (X_5), PDRB (X_1) pada titik 168,129.00, Tingkat partisipasi angkatan kerja (X_2) pada titik 64,53 dan titik 65,50, serta laju pertumbuhan penduduk (X_5)

pada titik 2,88 memenuhi semua asumsi identik, independen dan berdistribusi normal. Dengan nilai R^2 adalah 0,73 yang berarti kemampuan variabel respon dalam menjelaskan variabel prediktor sebesar 73% sedangkan sisanya dipengaruhi oleh variabel lain diluar penelitian. Sedangkan nilai MSE sebesar 7,786

Kesimpulan

Dari hasil analisis dan pembahasan yang telah dilakukan, maka dapat disimpulkan model terbaik dari IHK di Indonesia adalah pada tiga titik knot dengan nilai GCV minimum 14,07 maka diperoleh model regresi nonparametrik spline IHK di Indonesia adalah sebagai berikut:

$$\hat{y} = 6,81100 + 0,00010(X_1 - 168,129)^3 + (-0,12830)X_2 + 3,14100(X_2 - 64,53)^3 + (-3,76000)(X_2 - 65,51)^3 + 0,30850X_3 + 1,57200X_5 + (-8,01900)(X_5 - 2,88)$$

Model regresi nonparametrik spline tersebut memiliki empat variabel yang signifikan yaitu Tingkat partisipasi angkatan kerja (X_2), Bahan makanan (X_3), laju pertumbuhan penduduk (X_5).

Ucapan Terimakasih. Penulis mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah berkontribusi dalam penelitian ini.

Daftar Pustaka

- Budiantara, I. N. 2009. Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang, Pidato Pengukuhan Untuk Jabatan Guru Besar pada Jurusan Statistika, Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Surabaya: ITS Press.
- Budiantara, I.N. 2005. *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik*. Surabaya: Berkala MIPA, Institut Teknologi Sepuluh Nopember.
- Drapper, N.R., dan Smith, H. 1992. Analisis Regresi Terapan. Edisi Kedua. PT Gramedia Pustaka Utama. Jakarta.
- Eubank. 1998. Penentuan Generalized Cross Validation (GCV) Sebagai Kriteria dalam Pemilihan Model Regresi B-Spline Terbaik. *Jurnal Statistika*, 2(2), 121-126
- Gujarati, D. N. 2003. *Basic Econometrics Fourth Edition*. The McGraw-Hill Companies, Inc. 1221 Avenue of the Americas, New York.

Putri, T.D.P. 2018. *Pemodelan Regresi Nonparametrik Spline Multivariabel*. Semarang: Jurusan Matematika FMIPA Universitas Negeri Semarang.

Soleh, Achmad S. 2005. *Ilmu Statistika*. Penerbit Rekayasa Sains Bandung; Bandung.

Winarti & Sony, S. 2010. Pendekatan Regresi Semiparametrik Spline. (Pada Data Nilai Ujian Nasional Siswa SMKN 1 Nguling Pasuruan. *Jurnal Sains dan Seni Pomits*, 3(2), 194199.

Diterima pada tanggal 30 Januari 2022.

Terbit online pada tanggal 21 April 2022