

PEMODELAN PENDAPATAN ASLI DAERAH MENGGUNAKAN REGRESI NONPARAMETRIK LOKAL LINIER KERNEL

¹Martina Navra Tilova, ²Tiani Wahyu Utami, ³Moh. Yamin Darsyah

^{1,2,3}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang

Alamat e-mail : martinanavra48@gmail.com

ABSTRAK

Analisis regresi merupakan suatu metode statistika yang dapat digunakan untuk mengetahui hubungan antara suatu variabel terikat (dependen) Y terhadap satu atau lebih variabel bebas (independen) X sehingga memperoleh persamaan dan menggunakan persamaan tersebut untuk membuat perkiraan atau prediksi. Regresi nonparametrik adalah pendekatan regresi yang tidak diketahui bentuk kurva regresi dan tidak memberikan asumsi bentuk kurva regresi tertentu. Estimator yang digunakan dalam penelitian ini adalah estimator lokal linier yang mempunyai kelebihan yaitu mengestimasi fungsi disetiap titik sehingga model yang didapatkan lebih mendekati dengan pola data yang sesungguhnya. Estimator lokal linier diperoleh dengan metode WLS (*Weighted Least Square*), selain itu juga estimator ini sangat bergantung pada *bandwidth*(h) optimal yang ditentukan berdasarkan metode GCV (*Generalized Cross Validation*). Selanjutnya model regresi nonparametrik berdasarkan estimator lokal linier kernel diterapkan pada data Pendapatan Asli Daerah (PAD) di Provinsi Jawa Tengah dengan tiga variabel prediktor yaitu Jumlah Penduduk, Inflasi, dan Pengeluaran Daerah. Dalam penelitian ini akan menerapkan model regresi nonparametrik berdasarkan estimator lokal linier kernel. Output yang dihasilkan dihasilkan dari program R menunjukkan bahwa nilai *bandwidth* optimal sebesar 1.7 dengan nilai GCV minimum sebesar 0.9007611 dan nilai MSE yang diperoleh adalah sebesar 8.767846.

Kata Kunci : PAD, Regresi nonparametrik, Lokal linier, Kernel.

ABSTRACT

*Regression analysis is a statistical method that can be used to determine the relationship between a dependent variable Y to one or more independent variables X to obtain equations and use these equations to make predictions or predictions. Nonparametric regression is a regression approach that is not known to form a regression curve and does not provide an assumption in the form of a particular regression curve. The estimator used in this study is a linear local estimator which has the advantage of estimating functions at each point so that the model obtained is closer to the actual data pattern. The linear local estimator is obtained by the WLS (*Weighted Least Square*) method, besides that this estimator is very dependent on the optimal bandwidth (h) which is determined based on the Generalized Cross Validation (GCV) method. Furthermore, the nonparametric regression model based on a linear local kernel estimator was applied to data on Regional Original Revenue (PAD) in Central Java Province with three predictor variables namely Population, Inflation, and Regional Expenditure. In this study we will apply a nonparametric regression model based on a linear local kernel estimator.*

The output produced from the R program shows that the optimal bandwidth value is 1.7 with a minimum GCV value of 0.9007611 and the MSE value obtained is 8.767846.

Keywords: PAD, Nonparametric Regression, Linear Local, Kernel.

PENDAHULUAN

Analisis regresi merupakan suatu metode statistika yang dapat digunakan untuk mengetahui hubungan antara suatu variabel terikat (dependen) Y terhadap satu atau lebih variabel bebas (independen) X. Dalam analisis regresi terdapat dua pendekatan yang dapat digunakan untuk mengestimasi kurva regresi, yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik.

Pendekatan parametrik, bentuk hubungan antara variabel respon dengan variabel prediktor diketahui atau diperkirakan dari bentuk kurva regresi sedangkan regresi nonparametrik dilakukan jika bentuk kurva regresinya tidak diketahui. Kurva regresi nonparametrik diasumsikan *smooth*, Pendekatan regresi nonparametrik digunakan untuk mengestimasi kurva regresi serta metode ini lebih fleksibel untuk mensubstitusi nilai-nilai yang hilang antara variabel prediktor yang berdekatan.

Diberikan (y_i, x_i) yang memenuhi regresi nonparametrik, dengan y_i sebagai variabel respon dan x_i sebagai variabel prediktor. Hubungan y_i dengan x_i tidak diketahui fungsinya. Maka dapat dimodelkan hubungan antara y_i dengan x_i dapat dimodelkan dengan menggunakan model regresi nonparametrik, dengan mengestimasi parameter maka akan memperoleh estimator kurva regresinya dimana dalam tulisan ini akan digunakan estimator lokal linier kernel. Kelebihan antara lain Mengestimasi fungsi disetiap titik sehingga model yang dihasilkan mendekati dengan data sesungguhnya. Estimasi Lokal linier dapat

menggunakan WLS (*Weighted Least Square*) dengan cara meminimumkannya.

Penelitian terdahulu telah banyak yang dilakukan mengenai estimasi regresi nonparametrik berdasarkan estimator lokal linier kernel. *Generalized Exploratory Factor Analysis* (GEFA) dan estimator lokal linier multiprediktor dalam pemodelan kalibrasi senyawa aktif kurkumin.

Pendapatan Asli Daerah adalah pendapatan daerah yang dapat digunakan untuk masing-masing daerah yang menyelenggarakan pemerintah dan pembangunan daerah. Peningkatan PAD diharapkan meningkatkan investasi belanja modal pemerintah daerah sehingga kualitas pelayanan publik semakin baik, tetapi yang terjadi adalah peningkatan PAD tidak diikuti dengan kenaikan anggaran belanja modal yang signifikan.

Selanjutnya pada kesempatan ini akan dibahas bagaimana mengestimasi kurva regresi nonparametrik menggunakan estimator lokal linier kernel dan akan diterapkan pada data pendapatan asli daerah.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan adalah data sekunder yang diambil dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Dari data dapat diketahui Pendapatan Asli Daerah (PAD) sebagai variabel (Y), Jumlah Penduduk, Inflasi, dan Pengeluaran Daerah sebagai variabel (X) dari 35 kabupaten/kota

yang ada di Provinsi Jawa Tengah tahun 2016.

Metode Analisis

Langkah – langkah yang dilakukan sebagai berikut :

1. Mengestimasi model regresi nonparametrik berdasarkan estimator lokal linier kernel dengan langkah – langkah sebagai berikut:

- b. diberikan data observasi (y_i, x_i) , yang memenuhi regresi nonparametrik.

$$i. y_i = m(x_i) + \varepsilon ;$$

$$i = 1, 2, 3, \dots , n$$

- c. Menyatakan (y_i) dapat didekati oleh lokal linier $m(x_i) \approx m(x) + (x_i - x) m^{(1)}$

Jika $\beta_r(x) = \frac{m^{(r)}(x)}{r!}$ Maka

$$m(x_i) = \beta_r(x) + (x_i - x)\beta_1(x)$$

- d. Dengan α mengestimasi β dengan optimasi atau meminimumkan kriteria *Weighted Lest Square* :

$$M = \sum_{i=1}^n (y_i - m(x_i))^2 K_h(x_i - x)$$

Dengan $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$ fungsi pembobot kernel

- e. Menyatakan $\hat{m} \approx X\hat{\beta}$ dengan X adalah matrik berukuran $1 \times (p + 1)$ dan \hat{m} adalah vector berukuran 1×1

- f. Mendapatkan bentuk matrik $A(h)$ berukuran $N \times N$ dengan cara menyelesaikan persamaan berikut : $\hat{m} = A(h)y^*$

2. Langkah Pemodelan Pendapatan Asli Daerah dengan menggunakan Estimator Lokal Linier Kernel

Hasil estimasi yang diperoleh diterapkan untuk menganalisis pengaruh variabel pendapatan asli daerah terhadap

variabel jumlah penduduk, inflasi, dan pengeluaran daerah di Provinsi Jawa Tengah dengan langkah – langkah sebagai berikut :

- a. Membuat plot data berpasangan. $(y_i, x_i) \quad i = 1, 2, 3, \dots, 35.$
- b. Menentukan matriks $A(h)$ berukuran $N \times N.$
- c. Memilih nilai *bandwidth* optimal yang meminimumkan.

$$GCV(h) = \frac{MSE(h)}{\left(\frac{1}{n} \text{tr} [I - A(h)]\right)^2}$$

- d. Memodelkan orde lokal linier dan nilai *bandwidth* optimal dari langkah secara simultan.
- e. Menghitung nilai MSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

HASIL PENELITIAN

Berdasarkan tujuan dari penelitian maka pada bab ini akan dilakukan estimasi model regresi nonparametrik berdasarkan estimator lokal linier kernel dengan meminimumkan *Weighted Least Square* (WLS). Hasil estimasi yang diperoleh akan diterapkan pada model pendapatan asli daerah dengan menggunakan estimator lokal linier kernel.

Estimasi Model Regresi Nonparametrik Berdasarkan Estimator Lokal Linier Kernel

Diberikan data observasi (y_i, x_i) yang memenuhi regresi nonparametrik, dengan y_i sebagai variabel respond an x_i sebagai variabel prediktor. Hubungan y_i dengan x_i tidak diketahui fungsinya. Maka dapat dimodelkan hubungan antara y_i dengan x_i dapat dimodelkan dengan menggunakan model regresi nonparametrik sebagai berikut :

$$y_i = m(x_i) + \varepsilon ; \quad i = 1, 2, 3, \dots , n$$

Model dapat ditulis dalam bentuk matriks :

$$y = m + \varepsilon$$

dengan

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad m = \begin{bmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \text{dan} \quad = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

mengestimasi $m(x_i)$ didekatkan dengan polinomial derajat $p = 1$ sebagai berikut :

$$m(x_i) = m(x_i) + (x_i - x) m^{(1)}$$

Jika $\beta_r(x) = \frac{m^r(x)}{r!}$ Maka

$$m(x_i) = \beta_r(x) + (x_i - x)\beta_1(x_1) \quad (4.2)$$

Persamaan (4.2) dapat ditulis menjadi sebagai berikut :

$$m(x_i) = \hat{x}_i^T \beta \quad i = 1, 2, 3, \dots, n$$

dengan $\hat{x}_i = [1, (x_{i1} - x), (x_{i2} - x), (x_{i3} - x)]^T$ dan

$$\beta = [\beta_r, \beta_1]^T, \quad \beta_r(x) = \frac{m^r(x)}{r!}, \quad r = 0 \text{ dan } 1$$

fungsi $m(x_i)$ diperoleh dengan optimasi atau meminimumkan kriteria *Weighted Least Square* (WLS). Jika $S(\beta)$ merupakan fungsi WLS maka :

$$S(\beta) = (y - X\beta)^T K_h (y - X\beta)$$

Dengan $y_i = [y_1, y_2, \dots, y_n]$ $X = [x_1^T, \dots, x_n^T]$,
 $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$,

$$Y = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

K adalah fungsi kernel, h adalah *bandwidth*,
 $K_h = \text{diag} (K_{h1}, K_{h2}, \dots, K_{hn})$.

$$K_{hi} = \text{diag} (K_{hi}(x_1 - x), K_{hi}(x_2 - x), \dots, K_{hi}(x_n - x))$$

Nilai estimasi β yaitu $\hat{\beta}$ dapat diperoleh dengan cara meminimumkan WLS atau meminimumkan $S(\beta)$. Nilai minimum $S(\beta)$ didapat dengan pendiferensialan persamaan terhadap β atau $\frac{\partial S(\beta)}{\partial \beta} = 0$, sehingga

diperoleh :

$$\hat{\beta} = (X^T K_h X)^{-1} X^T K_h y$$

Dengan demikian $m = X\hat{\beta}$

$$= (X^T K_h X)^{-1} X^T K_h y$$

Pemodelan Pendapatan Asli Daerah dengan menggunakan Estimator Lokal Linier Kernel

Pemodelan regresi nonparametrik dengan menggunakan estimator lokal linier ini akan menggunakan data Pendapatan Asli Daerah Provinsi Jawa Tengah pada 35 Kabupaten/Kota. Pada penelitian ini akan dilihat hubungan antar variabel respons dan prediktor, dalam penelitian ini menggunakan beberapa variabel antara lain : Pendapatan Asli Daerah, Jumlah Penduduk, Inflasi dan Pengeluaran Daerah.

Estimasi model regresi nonparametrik lokal linier kernel sangat sensitive dalam pemilihan *Bandwidth*, dimana *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan fluktuatif. Sebaliknya, *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus. Oleh karena itu perlu

dipilih *bandwidth* yang optimal untuk menghasilkan kurva optimal. Nilai *bandwidth* dapat dilihat dari nilai GCV yang minimum dengan orde polinomial $p = 1$. Pemilihan ini digunakan untuk proses estimasi dengan fungsi kernel yang digunakan adalah fungsi Kernel Gaussian.

Pemilihan Nilai Bandwidth Optimal pada Model Regresi Nonparametrik

Langkah awal sebelum mengestimasi model regresi secara simultan adalah menentukan *bandwidth* optimal dan orde polinomial, karena nilai *bandwidth* menjadi salah satu faktor yang paling utama dalam penentuan estimasi parameter di regresi lokal linier. Selain *bandwidth* orde polinomial juga merupakan hal yang penting, dalam penelitian ini orde polinomial (p) telah ditentukan yaitu $p=1$ karena metode yang digunakan lokal linier. Nilai *bandwidth* yang optimal dengan orde polinomial ditentukan dengan metode GCV, dimana GCV ini merupakan salah satu metode untuk menentukan nilai *bandwidth* h yang optimal.

Tabel 1. Nilai GCV dan *Bandwidth* pada Pendapatan Asli Daerah, Jumlah Penduduk, Inflasi, dan Pengeluaran daerah

Orde Polinomial	<i>Bandwidth</i> (h)	GCV
1	1.0	0.9265110
1	1.1	0.9136619
1	1.2	0.9070739
1	1.3	0.9036786
1	1.4	0.9019601
1	1.5	0.9011433
1	1.6	0.9008180
1	1.7	0.9007611
1	1.8	0.9008481
1	1.9	0.9010090

Berdasarkan Tabel 1. dapat dilihat bahwa nilai GCV minimum adalah 0.9007611 dengan nilai *bandwidth* (h) sebesar 1.7 dengan demikian, model regresi nonparametrik dengan pendekatan lokal linier dengan nilai *bandwidth* sebesar 1.7 dengan orde polinomial $p=1$ nilai MSE sebesar 8.767846.

KESIMPULAN

berdasarkan pembahasan, maka dapat disimpulkan bahwa :

1. Estimasi model regresi nonparametrik dengan menggunakan estimator lokal linier kernel pada kasus pendapatan asli daerah yaitu sebagai berikut :

$$y = m + \varepsilon$$

$$\text{dengan } m = X\hat{\beta}$$

$$\hat{\beta} = (X^T K_h X)^{-1} X^T K_h y$$

2. Hasil penerapan model regresi nonparametrik pada data pendapatan asli daerah di Provinsi Jawa Tengah, maka diperoleh hasil pemodelan dengan nilai *bandwidth* optimal sebesar 1.7 dengan orde polinomial $p=1$ dan nilai MSE sebesar 8.767846. Pada *scatterplot* untuk semua pengamatan pendapatan asli daerah dengan jumlah penduduk, pendapatan asli daerah dengan inflasi, dan pendapatan asli daerah dengan pengeluaran daerah terlihat bahwa persebaran data tidak membentuk pola tertentu sehingga data tepat didekati dengan nonparametrik dan kurva yang dihasilkan dari nilai prediksi tidak mendekati

dengan nilai yang sebenarnya sehingga kemungkinan pendapatan Asli Daerah (PAD) ini dapat didekati dengan estimator lainnya.

Dalam penelitian ini masih memiliki banyak kekurangan karena dari permasalahan yang dikaji sendiri masih terbatas oleh karena itu perlu dilakukan penelitian selanjutnya dengan menggunakan estimator lain yang baik digunakan untuk memodelkan Pendapatan Asli Daerah ini Sendiri.

UCAPAN TERIMA KASIH

Pada penelitian ini saya ucapkan terima kasih kepada dosen pembimbing Tiani Wahyu Utami, M.Si dan Moh. Yamin Darsyah, M.Si yang telah banyak membantu dan membimbing dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] Abdul Halim. 2009. Akuntansi sector Publik Akuntansi Keuangan Daerah, Edisi Pertama. Jakarta: Salemba Empat
- [2] Budiantara, I.N. Lestari, B., dan Islamiyati, A., 2010. Estimator Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik Heterokedastisitas untuk Data Longitudinal, Hibah Penelitian Kompetensi, DP2M-Dikti: Jakarta.
- [3] Dewi, Elita. 2002. "Identifikasi Sumber Pendapatan Asli Daerah dalam Rangka Pelaksanaan Otonomi Daerah", Jurnal Ilmu Sosial dan Politik, Universitas Sumatera Utara.
- [4] Elita. 2007. Penerimaan Penerimaan Pendapatan Asli Daerah. Jakarta: Rajawali.
- [5] Eubank, R.L., Huang, C., Maldonado, Y.M., Wang, N., Wang, S., dan Buchanan, R.J., 2004, *Smoothing Spline Estimation in Varying-Coefficient Models*, J. R. Statist.Soc., 66, Part 3, pp. 653-667.
- [6] Gujarati, Damodar N. 2006. *Ekonometrika Dasar*. Jakarta : Penerbit Erlangga.
- [7] Hardle, Wolfgang. 1994. *Applied Nonparametric Regression*. Springer-Verlag. Berlin.
- [8] Latumaerissa, J.R. (2011). *Bank dan Lembaga Keuangan Lain*. Jakarta: Salemba empat.
- [9] Lauda mutia. 2016. *Generalized Exploratory Factor Analysis (GEFA) dan estimator lokal linier multiprediktor dalam pemodelan kalibrasi senyawa aktif kurkumin*. Universitas Airlangga
- [10] Mankiw, N. Gregory., Quah, Euston., dan Wilson, Peter. 2012. *Pengantar Ekonomi Makro*. Edisi Asia. Jakarta: Salemba Empat.
- [11] Utami, T. W., Nur, I. M., & Java, C. (2016). Longitudinal Data Modeling Based on Platelets Level in DHF (Dengue Hemorrhagic Fever) Using Nonparametric Regression of Local Polynomial Kernel GEE Approach, (November 2014), 0-3.