**ESTIMASI KURVA REGRESI NONPARAMETRIK DENGAN MENGGUNAKAN METODE FOURIER DENGAN DAN TANPA PENCILAN**

Neli Rohmatilah(1)\*, Netti Herawati(1), Dorrah Azis(2)

*(1)Jurusan Matematika FMIPA*

*Universitas Lampung, Bandar Lampung, 35145*

*\*email:* *neliromatilah.06@gmail.com**,* *netti.herawati@fmipa.unila.ac.id**,* *dorrah.azis@fmipa.unila.ac.id*

Diterima (Tgl Bulan Tahun), Direvisi (Tgl Bulan Tahun) (Times New Roman, 10pt)

***Abstract.*** *This study aims to estimate the nonparametric regression curve using Fourier method for data with and without outliers on the periodic and linear wave functions. The Simulation data is used and the best regression curve resulted from Fourier method is examined using Generalized. Cross Validation (GCV). The results showed that the estimation of nonparametric regression curves using the Fourier method on the periodic and linear wave functions is affected by outliers. The Fourier method estimates the best curve when the data does not contain outliers.*

***Keywords:*** *Fourier method, GCV, Nonparametric Regression*

**Abstrak.** Penelitian ini bertujuan untuk memperkirakan kurva regresi nonparametrik menggunakan metode Fourier untuk data dengan dan tanpa pencilan pada data fungsi gelombang periodik dan linear. Data simulasi digunakan dan kurva regresi terbaik yang dihasilkan dari metode Fourier diperiksa menggunakan *Generalized. Cross Validation (GCV).* Hasil penelitian menunjukkan bahwa estimasi kurva regresi nonprametrik menggunakan metode Fourier pada fungsi gelombang periodik dan linear dipengaruhi oleh pencilan. Metode Fourier memperkirakan kurva terbaik ketika data tidak menganduk pencilan.

**Kata kunci**:GCV, Metode Fourier, Regresi Nonparametrik

**PENDAHULUAN**

Analisis regresi merupakan metode statistika untuk mengetahui hubungan antara satu variabel respon (*dependen*) dengan satu atau lebih variabel prediktor (*independen*). Analisis regresi adalah hubungan antara variabel respon (*dependen*) dengan satu atau lebih variabel prediktor (*independen)* [4].

Pendekatan regresi dapat dilakukan dengan pendekatan parametrik dan nonparametrik. Pendekatan parametrik dilakukan apabila bentuk kurva regresi diketahui, dalam regresi parametrik data harus memenuhi asumsi normalitas. Sedangkan pendekatan nonparametrik dilakukan apabila bentuk kurva tidak diketahui. regresi nonparametrik tidak memerlukan asumsi-asumsi yang terikat seperti pendekatan parametrik, pendekatan nonparametrik mempunyai fleksibelitas yang tinggi karena tidak mengasumsikan bentuk kurva regresi.

Pada penelitian ini akan memperkirakan kurva regresi nonparametrik menggunakan metode Fourier untuk data dengan dan tanpa pencilan pada data fungsi gelombang periodik dan linear.

Kurva regresi diasumsikan termuat dalam ruang fungsi tertentu [3]. Secara umum model regresi nonparametrik adalah:

 $y\_{i}=f(x\_{i})+ε\_{i} i=1,2,3,…,n$ (1)

dengan $y\_{i}$ adalah variabel tak bebas pada pengamatan ke-i $f(x\_{i}) $merupakan kurva fungsi regresi yang tidak diketahui bentuknya dengan $x\_{i}$ merupakan variabel indepenen, $ε\_{i}$ merupakan galat yang berdistribusi normal, mean 0 dan varian $σ^{2} (0, σ^{2} )$. Dalam hal ini, $f(x\_{i})$ hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, dimana pemilihan ruang fungsi tersebut biasanya termotivasi oleh sifat kemulusan yang dimiliki oleh fungsi $f(x\_{i})$ tersebut [7].

Estimasi fungsi regresi nonparametrik dilakukan berdasarkan data pengamatan dengan menggunakan teknik pemulusan. Terdapat beberapa teknik pemulusan dengan model regresi nonparametrik antara lain penduga kernel, penduga spline, fourier dan wavelet [6]. Penelitian dengan menggunakan metode Fourier untuk nonparametrik telah banyak dikembangkan salah satunya adalah [5] penduga kurva regresi nonparametrik linear dan nonlinear dengan metode Fourier dan metode Nadaraya-Watson.

Metode Fourier adalah fungsi trigonometri yang mempunyai tingkat fleksibelitas. Hal ini karena Fourier merupakan kurva yang menunjukan fungsi sinus dan cosines. Jika fungsi f(x) terdefinisi pada interval $[-L, L ]$ dan diluar selang ini oleh $f\left(x\pm 2L\right)=f(x)$, sehingga $f(x)$ merupakan fungsi periodik dengan periode 2L [7]. $f\left(x\right)$ dapat direpresentasikan dengan deret perluasan Fourier sebagai berikut :

$f\left(x\right)=\frac{1}{2}a\_{0}+\sum\_{j=1}^{\infty }a\_{j}\cos(\left(\frac{2πjx}{2L}\right))+b\_{j}\sin(\left(\frac{2πjx}{2L}\right)) $ (2)

Nilai $\frac{2π}{T}$ (dengan T adalah periode $f\left(x\right)$) merupakan faktor pengali agar x dalam satuan radian.

Dari pendekatan deret Fourier yang digunakan terdapat estimator fourier diberikan n data pengamatan $\{\left(x\_{i},y\_{i}\right)\}\_{i=1}^{n}$ yang memenuhi **“persamaan 2”** yang didefinisikan sebagai berikut:

$\hat{m}\left(j,x\_{i}\right)=\frac{1}{2}a\_{0}+\sum\_{j=1}^{J}a\_{j}\cos(\left(wjx\_{i}\right))+b\_{j}\sin(\left(wjx\_{i}\right)) $ $w=\frac{2π}{2L}$ (3)

dengan $a\_{0}, a\_{j}$ dan $b\_{j}$ adalah koefisien Fourier.

Tingkat kemulusan estimator Fourier ditentukan oleh pemilihan parameter pemulus. Semakin kecil parameter pemulusan, semakin mulus estimasinya dan sebaliknya. Oleh karena itu, perlu dipilih *bandwidth* yang optimal. Salah satu metode yang dapat digunakan adalah metode *Generalized Cross Validation (GCV).* Penentuan *bandwidth* optimal akan menghasilkan nilai koefisien determinasi $\left(R^{2}\right)$ yang tinggi. *Generalized Cross Validation(GCV)* didefiniskan sebagai berikut:

$GCV \left(J\right)=\frac{n^{-1}\sum\_{i=1}^{n}(y\_{i}-\hat{m\_{f}}(J,x\_{i}))^{2}}{[1-(\frac{tra,e(A(A^{t}A)^{-1}A^{t})}{n})]^{2}}$ (4)

Nilai GCV terkecil akan menghasilkan nilai h yang optimal [2].

Selanjutnya kriteria untuk mentukan estimator terbaik dalam model regresi antara lain nilai *Mean Square Error (MSE)* dan nilai koefisien determinasi *R-Square* (R2). Kebaikan suatu penduga dapat dilihat dari tingkat kesalahannya. Semakin kecil tingkat kesalahan suatu pendugaan maka semakin baik estimasinya. [1] MSE didefinisikan sebagai berikut:

$MSE=\frac{1}{n}\sum\_{i=1}^{n}(y\_{i}-\hat{y\_{i }})^{2}$ (5)

sedangkan koefisen determinasi didefinisikan sebagi berikut:

$R^{2}=\frac{SSR}{SST}=\frac{\sum\_{i=1}^{n}(\hat{y\_{i}}-\overbar{y })^{2}}{\sum\_{i=1}^{n}(y\_{i}-\overbar{y})^{2}}$ (6)

**METODE PENELITIAN**

Data yang digunakan dalam penelitian ini adalah data simulasi menggunakan *software* Matlab 2013bdengan variabel bebas $x\~U(0,3)$ dengan n= 100 yang dipetakan pada data variabel respon $(y\_{i})$ melalui persamaan:

$y\_{i}=\sin(\left(4x\_{i}\right))+ε\_{I}$

$y\_{i}=\cos(\left(4x\_{i}\right))+ε\_{i}$

$y\_{i}=2x\_{i}+1+ε\_{i}$

dengan $ε\_{i} $merupakan galat dengan distribusi $ε\_{i}\~N\left(0,1\right),$ $ε\_{i}\~N\left(0,0.5\right)$ dan

$ ε\_{i}\~0.9N\left(0,1\right)+0.1N(5,15)$.

Adapun langkah-langkah pada penelitian ini sebagai berikut:

1. Menentukan nilai *bandwidth* optimal berdasarkan nilai GCV dan MSE minimum.
2. Membandingkan kurva estimasi regresi nonparametrik dengan metode Fourier pada fungsi.
3. Menguji kebaikan model regresi dengan menghitung R2.

**HASIL DAN PEMBAHASAN**

Pemilihan *bandwidth* yang optimal menggunakan metode *Generalized Cross Validation* (GCV) pada fungsi $y=\sin(\left(4x\right)+ε\_{i})$ dapat dilihat pada tabel 1.

**Tabel 1.** Nilai *Generalized Cross Validation (GCV)*

 untuk *bandwidth* optimal $y=\sin(\left(4x\right))+ε\_{i}$

|  |  |  |
| --- | --- | --- |
| Galat | *Bandwidth (h)* | GCV |
| $$N\~\left(0,1\right)$$ | 1.0 | 0.8979 |
| $$N\~\left(0,0.5\right)$$ | 1.0 | 0.2597 |
| $Normal (0.9N\left(0,5\right)+0.1$ $N(5,15)$ | 3.0 | 29.2775 |

Tabel 1 memperlihatkan nilai *Generalized Cross Validation* (GCV) minimum dengan metode Fourier dengan dan tanpa pencilan. Estimasi kurva regresi nonparametrik dengan metode Fourier pada bentuk gelombang periodik sinus dengan *bandwidth* optimal dapat dilihat pada Gambar 1-3.



**Gambar 1.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y\_{i}=\sin(\left(4x\_{i}\right))+ε\_{1}\~N(0,1)$

 dengan nilai *bandwidth* optimal (h= 1.0)



**Gambar 2.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=\sin(\left(4x\right))+$

$ ε\_{2}\~N(0,0.5)$ dengan *bandwidth* optimal

$ $(h= 1.0)



**Gambar 3.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=\sin(\left(4x\right))+ε\_{3}\~0.9N\left(0,5\right)$

 $+0,1N(5,15)$dengan *bandwidth* optimal

 (*h*= 3.0)

Dari Gambar 1 menunjukkan kurva regresi yang mendekati atau mengestimasi fungsi aslinya, pada Gambar 2 mengasilkan kurva regresi yang mendekati atau mengestimasi fungsi aslinya dan sesuai dengan pola data. Sedangkan pada gambar 3 memperlihatkan sebaran data yang terdapat pencilan 10% .

Ukuran kebaikan penduga juga dapat dapat dilihat dari nilai *Mean Square Error (MSE)* dan R-square untuk nilai *bandwidth* yang optimal seperti pada Tabel 2.

**Tabel 2.** Ukuran kebaikan metode Fourier dengan

 *bandwidth* optimal pada fungsi

 $y=\sin(\left(4x\right))+ ε\_{i}.$

|  |  |  |
| --- | --- | --- |
| Galat | MSE | R-Square |
| $$N\~\left(0,1\right)$$ | 0.8449 | 0.5100 |
| $$N\~\left(0,0.5\right)$$ | 0.2444 | 0.7113 |
| $Normal (0.9N\left(0,5\right)+0.1$ $N(5,15)$ | 26.9105 | 0.0002 |

Pada Tabel 2 memperlihatkan bahwa metode Fourier pada fungsi

$y=\sin(\left(4x\right)+ε)\~N\left(0,0.5\right) $memiliki MSE yang lebih kecil dari galat yang lainnya yaitu 0.2444 dengan nilai koefisien determinasi sebesar 71.13% yang artinya keragaman dari variabel respon dapat dijelaskan oleh variabel prediktor. Sehingga,, metode Fourier dengan fungsi $y=\sin(\left(4x\right))+ε\~N(0.05)$ dapat menduga kurva regresi nonparametrik pada fungsi gelombang periodik.

Pemilihan *bandwidth* yang optimal menggunakan metode *Generalized Cross Validation* (GCV) pada fungsi $y=\cos((4x)+)ε\_{i}$ dapat diihat pada tabel 3.

**Tabel 3.** Nilai *Generalized Cross Validation (GCV)*

untuk *bandwidth* optimal$y=\cos(\left(4x\right))+ε\_{i}$

|  |  |  |
| --- | --- | --- |
| Galat | *Bandwidth (h)* | GCV |
| $$N\~\left(0,1\right)$$ | 3.0 | 0.8360 |
| $$N\~\left(0,0.5\right)$$ | 1.0 | 0.2633 |
| $Normal (0.9N\left(0,5\right)+0.1$ $N(5,15)$ | 4.0 | 29.6995 |

Tabel 3 memperlihatkan nilai *Generalized Cross Validation* (GCV) minimum dengan metode Fourier dengan dan tanpa pencilan.

Penduga kurva regresi nonparametrik dengan metode Fourier pada bentuk gelombang periodik cosinus dengan *bandwidth* optimal dapat dilihat pada Gambar 4-6.



**Gambar 4.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=cos\left(4x\right)+ε\_{1}\~N(0,1)$

 dengan *bandwidth* optimal (*h*= 3.0)



**Gambar 5.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=cos\left(4x\right)+$

$ ε\_{2}\~N(0,0.5)$ dengan *bandwidth* optimal

 (*h*= 1.0)



**Gambar 6.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=cos\left(4x\right)+ε\_{3}\~0.9N\left(0,5\right)$

 $+0,1N(5,15)$ dengan*bandwidth* optimal

 (*h*= 4.0)

Dari Gambar 4 menunjukkan kurva regresi yang mendekati atau mengestimasi fungsi aslinya, pada gambar 5 mengasilkan kurva regresi yang mendekati atau mengestimasi fungsi aslinya dan sesuai dengan pola data. Sedangkan pada Gambar 6 memperlihatkan sebaran data yang terdapat pencilan 10% .

Ukuran kebaikan penduga juga dapat dapat dilihat dari nilai *Mean Square Error (MSE)* dan R-square untuk nilai *bandwidth* yang optimal seperti pada tabel 4.

**Tabel 4**. Ukuran kebaikan metode Fourier dengan

 *bandwidth* optimal pada fungsi

 $y=\cos(\left(4x\right))+ε\_{i}.$

|  |  |  |
| --- | --- | --- |
| Galat | MSE | R-Square |
| $$N\~\left(0,1\right)$$ | 0.7231 | 0.3550 |
| $$N\~\left(0,0.5\right)$$ | 0.2481 | 0.6270 |
| $$Normal(0.9N\left(0,5\right)+0.1N(5,15)$$ | 24.5941 | 0.0002 |

Pada Tabel 4 memperlihatkan bahwa metode Fourier pada fungsi $y=\cos(\left(4x\right)+ε)\~N(0,0.5)$ memiliki MSE yang lebih kecil dari galat yang lainnya adalah 0.2481 dengan nilai koefisien determinasi sebesar 62.70% yang artinya keragaman dari \_onparam respon dapat dijelaskan oleh \_onparam \_onparame. Sehingga, metode Fourier dengan fungsi $y=\cos(\left(4x\right))+ε\~N(0.05)$ dapat menduga kurva regresi nonparametrik pada fungsi gelombang perodik.

Pemilihan *bandwidth* yang optimal menggunakan metode *Generalized Cross Validation* (GCV) pada fungsi $y=2x+1+ε\_{i}$ dapat diihat pada tabel 5.

**Tabel 5.** Nilai *Generalized Cross Validation (GCV)*

untuk *bandwidth* optimal $y=2x+1+ε\_{i}$

|  |  |  |
| --- | --- | --- |
| Galat | *Bandwidth (h)* | GCV |
| $$N\~\left(0,1\right)$$ | 1.0 | 0.8831 |
| $$N\~\left(0,0.5\right)$$ | 2.0 | 0.2679 |
| $Normal (0.9N\left(0,5\right)+0.1$ $N(5,15)$ | 4.0 | 28.4911 |

Tabel 4 memperlihatkan nilai *Generalized Cross Validation* (GCV) minimum dengan metode Fourier dengan dan tanpa pencilan.

Penduga kurva regresi \_onparametric dengan metode Fourier pada fungsi linear dengan *bandwidth* optimal dapat dilihat pada Gambar 7-9.



**Gambar 7.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=2x+1+$ $ε\_{1}\~N(0,1)$

 dengan *bandwidth* optimal (*h*=1.0)



**Gambar 8.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=2x+1+ε\_{2}\~N(0,0.5)$

 dengan *bandwidth* optimal (*h*=2.0)



**Gambar 9.** Kurva Estimasi dengan metode Fourier

 pada fungsi $y=2x+1+$ $ε\_{3}\~0.9N\left(0,5\right)$

 $+0,1N\left(5,15\right)$dengan *bandwidth* optimal

 (*h*=4.0)

Dari Gambar 7 menunjukkan kurva regresi yang mendekati atau mengestimasi fungsi aslinya, pada gambar 8 mengasilkan kurva regresi yang mendekati atau mengestimasi fungsi aslinya dan sesuai dengan pola data. Sedangkan pada Gambar 9 memperlihatkan sebaran data yang terdapat pencilan 10% .

Ukuran kebaikan penduga ukuran kebaikan penduga juga dapat dapt dilihat dari nilai *MSE* dan R-square untuk nilai *bandwidth* yang optimal seperti pada Tabel 6.

**Tabel 6.** Ukuran kebaikan metode Fourier dengan

 *bandwidth* optimal pada fungsi

 $y=2x+1+ε\_{i}.$

|  |  |  |
| --- | --- | --- |
| Galat | MSE | R-Square |
| $$N\~\left(0,1\right)$$ | 0.8310 | 0.7231 |
| $$N\~\left(0,0.5\right)$$ | 0.2418 | 0.9196 |
| $$Normal(0.9N\left(0,5\right)+0.1N(5,15)$$ | 23.5936 | 0.0002 |

Pada Tabel 6 memperlihatkan bahwa metode Fourier pada fungsi $y=2x+1+ε\~N(0,0.5)$ memiliki MSE yang lebih kecil dari galat yang lainnya yaitu 0.2481 dengan nilai koefisien determinasi sebesar 91.96% yang artinya keragaman dari variabel respon dapat dijelaskan oleh variabel prediktor. Sehingga, dapat disimpulkan bahwa metode Fourier dengan fungsi $y=2x+1+ε\~N(0.05)$ dapat menduga kurva regresi nonparametrik pada berbentuk linear.

**KESIMPULAN**

Disimpulkan bahwa estimasi kurva regresi nonparametrik menggunakan metode Fourier pada fungsi gelombang periodik dan linear sangat baik menduga kurva regresi pada galat kecil dan tidak mengandung pencilan.

**DAFTAR PUSTAKA**

1. Chaterjee, S. 2007. *Regression Analysis by Example.* 4th Edition. Jhon Wiley and Sons, New Jersey.
2. Craven, P. and Wahba, G. 1979. Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smooting by The Method of Generalized Cross Validation.  *Numer Math University of Wisconsin*. **31**:377-403.
3. Eubank, R.L. 1998. *Spline Smoothing and Nonparametrik Regression.* Marcel Dekker, New York.
4. Hosmer, D.W. and Lemeshow, S. 2000. *Applied Logistic Regression*, John Wiley and Sons, New York.
5. Rudianto, J. 2015. Penduga Kurva Regresi Nonparametrik Linear dan Nonlinear dengan Metode Fourier dan Metode Nadaraya-Watson. Skripsi. Jurusan Matematika FMIPA Universitas Lampung, Bandar Lampung.
6. Silverman, B.W.  *1986. Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
7. Silverman, B.W. and Green, P.J. 1994. *Nonparametric Rgression and Generelized Linear Models*. Chapman & Hall/CRC, London.
8. Tolstov, G.P. 1962. *Fourier Series Translated from the Russian by Richard A. Silverman.* Dover Publications, New York.