

STUDI SIMULASI DAN APLIKASI: ESTIMATOR DERET FOURIER PADA PEMODELAN REGRESI NONPARAMETRIK

Andrea Tri Rian Dani^{1*}, Ardiana Fatma Dewi¹, Ludia Ni'matuzzahroh¹

¹Departemen Statistika, Fakultas Sains dan Analitika Data,
Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

Corresponding author: andreatriandani98@gmail.com

Abstrak. Diberikan data berpasangan t_i dan y_i , dimana $i=1,2, \dots, n$, selanjutnya hubungan antara variabel prediktor (t_i) dan respon (y_i) dapat dituliskan dalam bentuk fungsional sebagai berikut: $y_i = g(t_i) + \varepsilon_i$. Kurva regresi dari $g(t_i)$ tidak diketahui bentuknya, dan diasumsikan *smooth*, sehingga akan dihipotesiskan menggunakan estimator deret Fourier. Estimator deret Fourier merupakan polinomial trigonometri yang cenderung memiliki fleksibilitas tinggi, sehingga mampu menyesuaikan terhadap sifat lokal data. Tujuan dari penelitian ini adalah memodelkan data simulasi dan aplikasi menggunakan estimator deret Fourier. Studi simulasi dilakukan dengan membangkitkan fungsi trigonometri dan *error* yang mengikuti distribusi Normal. Data aplikasi pada penelitian ini adalah Rata-Rata Lama Sekolah di Provinsi Nusa Tenggara Timur (NTT) Tahun 2020. Metode estimasi yang digunakan adalah *Ordinary Least Squares* (OLS), selanjutnya metode yang akan digunakan dalam proses penentuan banyaknya osilasi yang optimal menggunakan *Generalized Cross-Validation* (GCV). Kriteria kebaikan model yang digunakan adalah Koefisien Determinasi (R^2). Berdasarkan hasil studi simulasi, diketahui jika estimator deret Fourier sangat dipengaruhi oleh banyaknya osilasi, apabila jumlah osilasi yang digunakan tepat dan optimal, maka diperoleh hasil estimasi kurva regresi yang baik. Selanjutnya berdasarkan kajian aplikasi, diperoleh model regresi nonparametrik deret Fourier yang terbaik dengan banyaknya osilasi adalah 4, yang ditunjukkan dengan nilai GCV yang paling minimum yaitu sebesar 0,27.

Kata Kunci: *Deret Fourier, Generalized Cross-Validation, Regresi Nonparametrik*

1 PENDAHULUAN

Analisis regresi merupakan salah satu metode statistika yang dapat digunakan untuk mengetahui pola hubungan antara variabel respon (Y) dan variabel prediktor (X), dengan tujuan utama yaitu untuk mencari bentuk estimasi kurva regresi [1]. Estimasi kurva regresi dapat dilakukan dengan beberapa pendekatan diantaranya yaitu pendekatan parametrik dan nonparametrik. Deteksi awal untuk mengetahui pola hubungan antara variabel respon dan prediktor menggunakan diagram pencar.

Pendekatan regresi parametrik dapat digunakan ketika kurva regresi diketahui bentuk pola hubungannya secara tertentu seperti linier, kuadrat, kubik, polinomial derajat p dan lain sebagainya [2]. Pada kenyataannya sekarang ini jarang ditemukan data yang mengikuti suatu pola tertentu, sehingga dibutuhkan pendekatan regresi nonparametrik. Pendekatan regresi nonparametrik dapat digunakan ketika bentuk kurva regresi diasumsikan tidak diketahui [3]. Terdapat beberapa estimator dalam regresi nonparametrik yang dikembangkan oleh para peneliti, diantaranya estimator Histogram, Spline, Kernel, Polinomial Lokal, dan Deret Fourier [4].

Salah satu estimator yang sering digunakan para peneliti akhir-akhir ini adalah Deret Fourier. Penelitian mengenai estimator regresi nonparametrik yang telah dilakukan sebelumnya dengan menggunakan estimator Deret Fourier oleh beberapa peneliti diantaranya [5], [6], [7], [8] dan [9]. Estimator Deret Fourier umumnya digunakan apabila data yang diselidiki polanya tidak diketahui dan terdapat kecenderungan pola berulang. Berdasarkan sifat fleksibilitas dan kelebihan estimator Deret Fourier, maka pada penelitian ini bertujuan untuk memodelkan data simulasi dan data aplikasi yaitu Rata-rata Lama Sekolah (RLS) di Kab/Kota Provinsi Nusa Tenggara Timur (NTT) pada Tahun 2020 menggunakan estimator Deret Fourier.

Pendidikan merupakan hal utama dalam pembangunan manusia dimana salah satu kualitas pendidikan di suatu wilayah dapat diketahui melalui Rata-rata Lama Sekolah (RLS). Beberapa faktor yang mempengaruhi diantaranya adalah faktor ekonomi berupa tingkat kemiskinan yang merupakan salah satu faktor penting penyebab putus sekolah. Rasio ketergantungan juga mempengaruhi pemerataan pendidikan, permintaan suatu keluarga untuk mendapatkan sejumlah anak ditentukan oleh preferensi keluarga itu sendiri atas jumlah anak yang dianggap bisa terus bertahan hidup. Hal lain yang dapat mempengaruhi rata-rata lama sekolah yaitu Jumlah Rumah Tangga Penerima Manfaat Program Indonesia Pintar (PIP), dimana dampak positif dari adanya PIP yaitu dapat meningkatkan angka partisipasi dan menurunnya jumlah anak putus sekolah.

Provinsi Nusa Tenggara Timur (NTT) merupakan salah satu Provinsi yang mutu pendidikannya menurun dan masih berada di peringkat bawah urutan nasional di Indonesia. Berdasarkan studi literasi yang dilakukan sebelumnya, maka pada penelitian ini akan difokuskan untuk studi simulasi dan aplikasi estimator Deret Fourier pada regresi nonparametrik yang diaplikasikan pada faktor-faktor yang mempengaruhi Rata-rata Lama Sekolah (RLS) di Provinsi Nusa Tenggara Timur (NTT) pada Tahun 2020.

2 TINJAUAN PUSTAKA

2.1 Regresi Nonparametrik

Regresi nonparametrik menunjukkan bentuk hubungan antara variabel respon dan variabel prediktor yang tidak diketahui bentuk polanya. Pendekatan regresi

nonparametrik memiliki fleksibilitas yang tinggi, karena data diharapkan mencari sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh faktor subjektivitas peneliti [4]. Terdapat sekumpulan data berpasangan (y_i, x_i) , dimana $i = 1, 2, \dots, n$, maka secara umum regresi nonparametrik dapat dituliskan sebagai berikut.

$$y_i = m(x_i) + \varepsilon_i \tag{1}$$

Dengan y_i merupakan variabel respon, m_i merupakan variabel prediktor, ε_i adalah *error* yang diasumsikan identik, independen, dan berdistribusi normal dengan mean nol dan varian σ^2 , sedangkan $m(x_i)$ merupakan fungsi regresi yang tidak diketahui bentuk polanya. Salah satu estimator yang dapat digunakan dalam regresi nonparametrik adalah Deret Fourier.

2.2 Estimator Deret Fourier

Estimator Deret Fourier pada regresi nonparametrik memiliki komponen prediktor yang bersifat aditif [5]. Estimator deret Fourier yang digunakan adalah jumlahan dari suatu fungsi linier dan suatu fungsi polinomial trigonometri. Deret Fourier umumnya digunakan apabila data yang diselidiki terdapat kecenderungan pola yang berulang. Deret Fourier merupakan polinomial trigonometri yang mempunyai fleksibilitas, sehingga dapat menyesuaikan diri secara efektif terhadap sifat lokal data, dimana pada estimasi kurva Deret Fourier dipengaruhi oleh panjang osilasi (w), semakin panjang osilasi maka gelombang yang dihasilkan semakin sering. Fungsi regresi $m(x_i)$ tidak diketahui secara pasti bentuk polanya dan diasumsikan termuat dalam ruang fungsi kontinu $C(0, \pi)$. Error random ε_i diasumsikan berdistribusi normal independen dengan mean 0 dan variansi σ^2 . Fungsi $m(x_i)$ membentuk pola hubungan yang cenderung berulang, maka dihipotesiskan dengan fungsi Deret Fourier berikut:

$$m(x_i) = \delta x_i + \frac{1}{2} \theta_0 + \sum_{w=1}^W \theta_w \cos wx_i \tag{2}$$

Dimana $\delta, \theta_0, \theta_w, w = 1, 2, \dots, W$ merupakan parameter-parameter dalam model.

2.3 Generalized Cross-Validation (GCV)

Salah satu metode yang sering digunakan dalam menentukan panjang osilasi (w) yang optimal pada Deret Fourier adalah *Generalized Cross Validation* (GCV) [10], [11]. Metode GCV mempunyai beberapa kelebihan dibandingkan dengan metode lainnya, seperti *Cross Validation* (CV) dan atau metode *Unbiased Risk* (UBR). Metode GCV secara teoritis memiliki sifat optimal asimtotik, formulanya tidak memuat varians σ^2 populasi yang tidak diketahui, serta *invariance* terhadap transformasi [10]. Fungsi GCV untuk pemilihan panjang osilasi optimal dapat ditunjukkan pada Persamaan (3).

$$GCV(w) = \frac{MSE(w)}{\left(n^{-1} \text{tr}(I - \mathbf{A}(w))\right)^2} \tag{3}$$

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

2.4 Koefisien Determinasi (R^2)

Salah satu tujuan analisis regresi adalah mendapatkan model terbaik yang mampu menjelaskan hubungan antara variabel prediktor dan variabel respon berdasarkan kriteria tertentu. Salah satu kriteria yang digunakan dalam pemilihan model terbaik adalah (R^2). Secara umum semakin besar nilai R^2 , maka semakin baik pula model yang didapatkan. Rumus untuk mencari koefisien determinasi didefinisikan sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

dengan y_i merupakan variabel respon ke- i , \hat{y}_i merupakan penduga variabel respon ke- i , dan \bar{y} merupakan rata-rata dari variabel respon

3 DATA DAN METODOLOGI PENELITIAN

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS) pada Tahun 2020 [12], dengan observasi penelitiannya adalah Kabupaten/Kota yang terdapat di Provinsi Nusa Tenggara Timur (NTT). Berikut merupakan definisi operasional masing-masing variabel yang digunakan dalam penelitian ini:

- Rata-rata Lama Sekolah (RLS) (Y)
Rata-rata jumlah tahun yang dihabiskan oleh penduduk berusia 15 tahun ke atas untuk menempuh semua jenis pendidikan yang pernah dijalani.
- Persentase Penduduk Miskin (X_1)
Persentase penduduk miskin yang berada di bawah garis kemiskinan.
- Rasio Ketergantungan (X_2)
Perbandingan antara jumlah penduduk umur 0-14 tahun, ditambah dengan jumlah penduduk 65 tahun ke atas dibandingkan dengan jumlah penduduk usia 15-64 tahun (angkatan kerja).
- Jumlah Rumah Tangga Penerima Manfaat Program Indonesia Pintar (PIP) (X_3)
Jumlah rumah tangga penerima manfaat Program Indonesia Pintar (PIP) menurut pihak yang mencairkan PIP dalam setahun terakhir dibagi Jumlah rumah tangga penerima manfaat Program Indonesia Pintar (PIP) dalam setahun terakhir dikali seratus.

1.2 Metodologi Penelitian

Berikut merupakan langkah-langkah atau tahapan pada penelitian ini:

- Melakukan studi simulasi dengan langkah-langkah sebagai berikut:
 - Ditulisiskan model regresi nonparametrik deret Fourier sebagai berikut:
$$y = m(x_i) + \varepsilon_i, \text{ dimana } i = 1, 2, \dots, n.$$

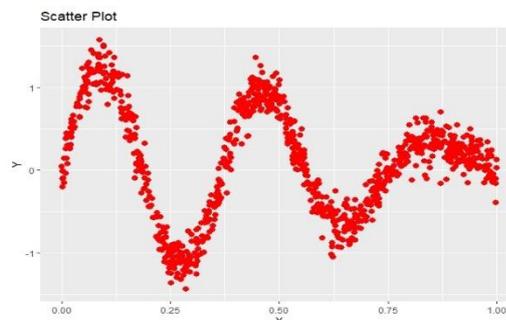
- b. Dibangkitkan data sebanyak 1000 ($n=1000$ dan $error$ yang mengikuti distribusi Normal $(0, \sigma^2)$). Pada penelitian ini σ^2 telah ditetapkan yaitu 0,15.
 - c. Menentukan bentuk kurva regresi dari fungsi trigonometri yaitu: $m(x_i) = 0,75 \sin(5\pi x_i) + 0,50 \sin(6\pi x_i)$.
 - d. Dibangkitkan variabel prediktor (x_i) yang saling independen dari distribusi Uniform $(0,1)$.
 - e. Berdasarkan Persamaan poin (a), maka akan diperoleh variabel respon (y_i).
 - f. Setelah didapatkan variabel respon (y_i) dan variabel prediktor, langkah selanjutnya membuat diagram pencar.
 - g. Memodelkan variabel respon (y_i) dengan variabel prediktor (x_i) dengan model regresi nonparametrik deret Fourier. Banyaknya osilasi yang akan dicobakan akan diduga dari pola yang terbentuk dari diagram pencar.
 - h. Menentukan banyaknya osilasi yang optimal menggunakan metode *Generalized Cross-Validation* (GCV). Nilai GCV yang minimum mengindikasikan jika banyaknya osilasi yang digunakan telah optimal.
2. Aplikasi estimator Deret Fourier pada Data Permasalahan Rata-Rata Lama Sekolah Kab/Kota di Provinsi Nusa Tenggara Timur pada Tahun 2020.
 - a. Melakukan analisis statistika deskriptif.
 - b. Membuat diagram pencar pada masing-masing variabel respon dan prediktor.
 - c. Pemodelan menggunakan pendekatan regresi nonparametrik estimator Deret Fourier.
 - d. Pemilihan model terbaik didasarkan pada nilai GCV yang minimum.

4. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan mengenai studi simulasi dan aplikasi estimator deret Fourier pada pemodelan regresi nonparametrik:

a. Studi Simulasi

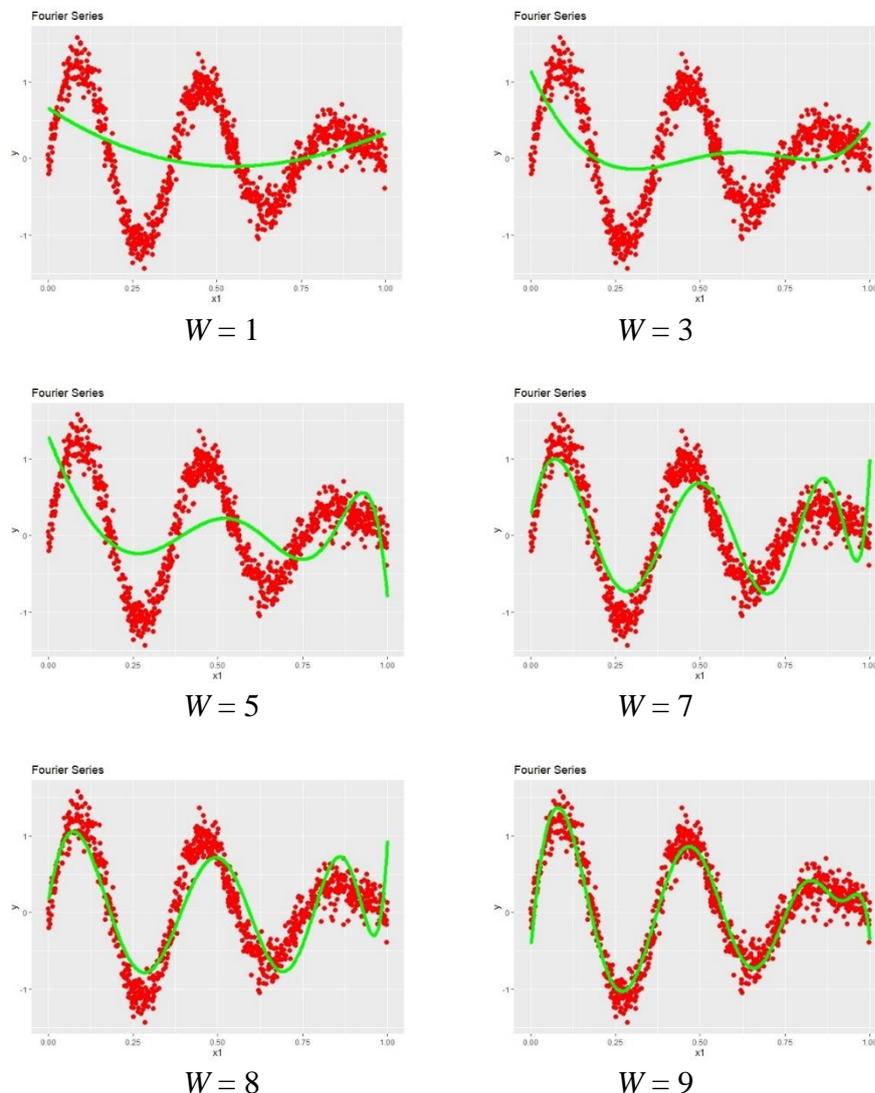
Pada kajian simulasi ini, akan dibangkitkan pola hubungan antara variabel prediktor (x_i) dengan variabel respon (y_i). Adapun diagram pencar yang terbentuk dari variabel respon (y_i) dengan variabel prediktor (x_i) disajikan pada Gambar 1.



Gambar 1: Diagram Pencar

Gambar 1 menunjukkan adanya pola yang berulang dan cenderung membentuk pola gelombang sinus-cosinus, sedemikian sehingga akan dilakukan pemodelan regresi nonparametrik deret Fourier untuk mengestimasi bentuk kurva regresinya. Metode estimasi yang digunakan yaitu *Ordinary Least Squares* (OLS).

Kelebihan pendekatan regresi nonparametrik khususnya estimator deret Fourier yaitu cenderung fleksibel dalam memodelkan pola data, sedemikian sehingga akan memberikan hasil yang tidak kaku dan lebih baik dibandingkan model regresi parametrik. Pada regresi nonparametrik deret Fourier terdapat yang namanya osilasi, dimana osilasi ini merupakan suatu parameter penghalus yang akan mengatur kemulusan kurva regresi. Osilasi inilah yang nantinya akan menangkap pola berulang dari data. Pada kajian simulasi ini, akan dicobakan beberapa variasi banyaknya osilasi, untuk menunjukkan dampak apabila osilasi yang digunakan tidak optimal.



Gambar 2: Hasil Estimasi Kurva Regresi dengan Variasi Nilai W

Nilai W yang optimal dipilih berdasarkan nilai GCV minimum. Hasil analisis untuk nilai W diberikan pada Tabel 1.

Tabel 1: Hasil Perhitungan Nilai GCV

Banyaknya Osilasi (W)	GCV
1	94299,48
2	40980,04
3	22189,03
4	12673,11
5	8386,71
6	8153,47
7	2480,95
8	2108,24
9	482,15

Berdasarkan Gambar 2, terlihat jika dengan $W = 9$ maka hasil estimasi kurva regresi cenderung mengikuti pola data aktualnya, dengan nilai GCV yang paling minimum yaitu sebesar 482,15.

b. Aplikasi pada Permasalahan Rata-Rata Lama Sekolah Kab/Kota di Provinsi NTT Tahun 2020

Pada bagian ini, akan dijabarkan mengenai statistika deskriptif, identifikasi awal bentuk pola data menggunakan diagram pencar, dan pemodelan regresi nonparametrik deret Fourier

1) Statistika Deskriptif

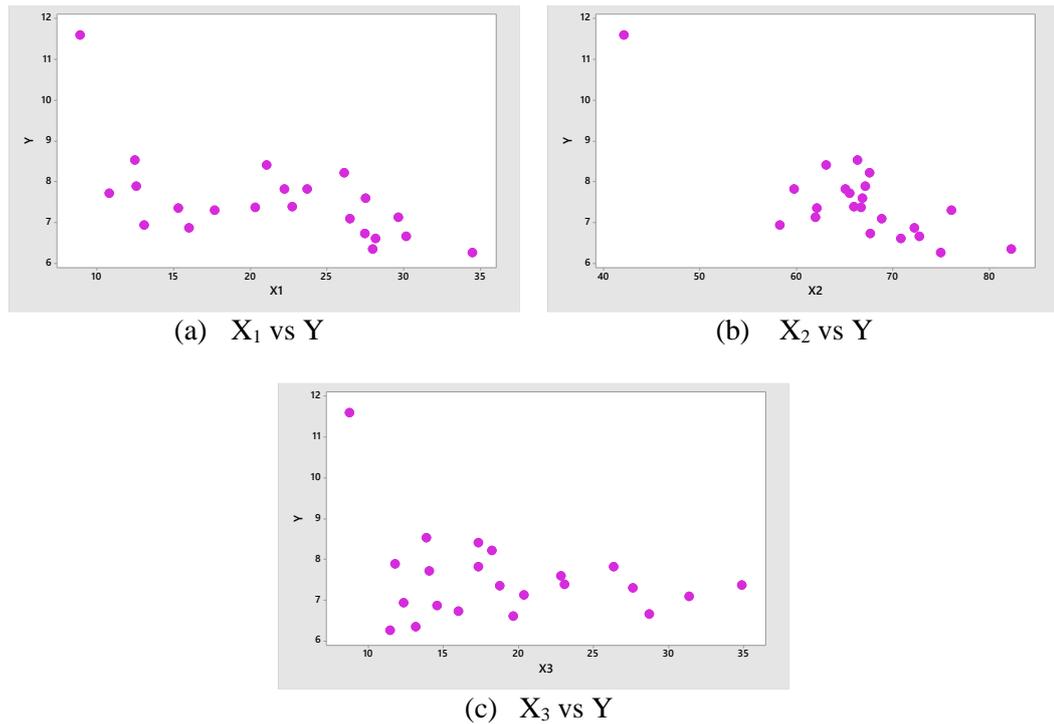
Sebelum dilakukan pemodelan regresi nonparametrik deret Fourier pada kasus rata-rata lama sekolah Kabupaten/Kota di Provinsi Nusa Tenggara Timur Tahun 2020, maka langkah awal yang perlu dilakukan adalah menampilkan statistika deskriptif guna memberikan informasi awal dari masing-masing variabel yang digunakan. Statistika deskriptif disajikan pada Tabel 2.

Tabel 2: Statistika Deskriptif

No.	Variabel	Nilai Minimum	Nilai Maksimum	Rata-Rata
1.	Rata-Rata Lama Sekolah	6,25	11,58	7,53
2.	Persentase Penduduk Miskin	8,96	34,49	21,62
3.	Angka Beban Ketergantungan	42,12	82,36	66,60
4.	Persentase Rumah Tangga yang Menerima Program Indonesia Pintar	8,78	34,91	19,24

2) Diagram Pencar

Untuk melakukan identifikasi awal mengenai pola hubungan antara setiap variabel prediktor terhadap variabel respon, dapat menggunakan diagram pencar. Adapun diagram pencar ditampilkan pada Gambar 3.



Gambar 3: Diagram Pencar

Pada Gambar 3, terlihat jika pola hubungan antara setiap variabel prediktor terhadap variabel respon cenderung tidak membentuk suatu pola tertentu, sehingga apabila digunakan model regresi parametrik akan kurang tepat. Pendekatan yang digunakan pada penelitian ini adalah regresi nonparametrik deret Fourier.

3) Pemodelan Regresi Nonparametrik Deret Fourier

Pada pemodelan regresi nonparametrik deret Fourier, langkah awal yang perlu diperhatikan adalah menentukan banyaknya osilasi yang digunakan. Banyaknya osilasi pada penelitian ini akan dibatasi, yaitu 1 sampai dengan 4 osilasi. Hasil pemodelan regresi nonparametrik deret Fourier ditampilkan pada Tabel 3.

Tabel 3: Nilai GCV Hasil Pemodelan dengan Deret Fourier

Banyaknya Osilasi (W)	GCV	R^2
1	4,69	69,98%
2	1,89	72,80%
3	0,91	76,54%
4	0,27	89,18%

Berdasarkan Tabel 3, diperoleh nilai GCV minimum yaitu saat banyaknya osilasi yang digunakan adalah 4 sebesar 0,27. Banyaknya osilasi yang dicobakan sebenarnya bisa lebih dari 4, namun hal ini akan berdampak langsung terhadap banyaknya parameter yang perlu diestimasi sehingga nantinya model yang dihasilkan tidak parsimoni. Berikut dituliskan model regresi nonparametrik deret Fourier dengan 4 osilasi sebagai berikut:

$$\begin{aligned}\hat{y}_i &= 18,27 - 0,05x_{1i} + 0,04\cos x_{1i} - 0,19\cos 2x_{1i} + 0,06\cos 3x_{1i} - 0,34\cos 4x_{1i} \\ &= -0,14x_{2i} + 0,25\cos x_{2i} + 0,71\cos 2x_{2i} - 0,09\cos 3x_{2i} - 0,10\cos 4x_{2i} + \\ &= 0,02x_{3i} + 0,26\cos x_{3i} + 0,03\cos 2x_{3i} - 0,65\cos 3x_{3i} - 0,92\cos 4x_{3i}\end{aligned}$$

Didapatkan nilai koefisien determinasi (R^2) sebesar 89,18%.

5. KESIMPULAN

Berdasarkan hasil analisis, maka dapat kita tuliskan kesimpulan sebagai berikut:

- 1) Berdasarkan hasil studi simulasi, diketahui jika estimator deret Fourier sangat baik dalam memodelkan pola data yang berulang dan yang memiliki pola seperti gelombang sinus-cosinus. Hasil estimasi kurva regresi cenderung mengikuti pola data aktualnya dengan syarat apabila banyaknya osilasi yang digunakan adalah tepat dan optimal.
- 2) Berdasarkan kajian aplikasi, diperoleh model regresi nonparametrik deret Fourier yang terbaik dengan banyaknya osilasi adalah 4, yang ditunjukkan dengan nilai GCV yang paling minimum yaitu sebesar 0,27.

Saran yang dapat diajukan untuk penelitian selanjutnya diantaranya:

- 1) Pada studi simulasi, peneliti tidak mempertimbangkan efek jika terdapat data pencilan, menjadi hal yang menarik apabila dilakukan kajian lanjutan mengenai hal ini.
- 2) Mengembangkan model estimator campuran untuk setiap variabel prediktor yang disesuaikan berdasarkan karakteristik pola hubungannya terhadap variabel respon.

DAFTAR PUSTAKA

- [1] Silverman, B. W. (1985). [Some Aspects of The Spline Smoothing Approach to Non-parametric Regression Curve Fitting](#). *Journal of the Royal Statistical Society. Series B (Methodological)*. Vol. 47, No. 1, hal: 01-52.
- [2] Budiantara, I. N. (2011). Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter. *Makalah Pembicara Utama pada Seminar Nasional FMIPA Undiksha, Bali*. hal: 09-28.
- [3] Lestari, B. & Budiantara, I. N. (2010). Spline Estimator of Triple Response Nonparametric Regression Model. *Jurnal Ilmu Dasar*, Vol. 11, hal: 17-22.
- [4] Eubank, R. L. 1999. *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.
- [5] Bilodeau, M. (1992). Fourier Smoother and Additive Models. *The Canadian Journal of Statistics*, Vol. 20, No. 03, hal: 257-259.
- [6] Tripena, A. & Budiantara, I. N. (2007). Fourier Estimator in Nonparametric Regression. *International Conference on Natural Sciences and Applied Natural Sciences*. Yogyakarta: Ahmad Dahlan University.
- [7] Asrini, L. J. & Budiantara, I. N. (2014). Fourier Series Semiparametric Regression Models (Case Study: The Production of Lowland Rice Irrigation in Central Java). *ARPN Journal of Engineering and Applied Sciences*. Vol. 09, No. 09, hal: 1501-1506.

- [8] Prahutama, A. (2013). Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur. *Prosiding Seminar Nasional Statistika Undip*. Vol. 10, hal: 69-76.
- [9] Adrianingsih, N. Y., Dani, A. T. R., & Ainurrochmah, A. (2020). Pemodelan Dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Nusa Tenggara Timur. *Jurnal EDUSAINTEK*. Vol. 4, hal: 400-407.
- [10] Wang, Y. (1998). Smoothing Spline Models with Correlated Random Errors. *Journal of the American Statistical Association*. Vol. 93, hal: 341-348.
- [11] Dani, A. T. R. & Adrianingsih, N. Y. (2020). Pemodelan Regresi Nonparametrik dengan Estimator Spline Truncated vs Deret Fourier. *Jambura Journal of Mathematics*. Vol. 03, No. 01, hal: 26-36.
- [12] Badan Pusat Statistik (BPS). (2020). *Provinsi Nusa Tenggara Timur dalam Angka 2020*. NTT: BPS.