

Nonparametric Regression Modeling with Fourier Series Approach on Poverty Cases in West Sumatra Province

Melin Wanike Ketrin, Fadhilah Fitri*, Atus Amadi Putra, Zilrahmi

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: fadhilahfitri@fmipa.unp.ac.id

Submitted : 18 Januari 2023
Revised : 02 Februari 2023
Accepted : 13 Februari 2023

ABSTRACT

Poverty is a complex problem that has an impact on various social problems such as education, unemployment, health and economic growth. Therefore, this is important to triumph over on to create population welfare. Regression analysis is one of the analyses that can be used to predict. Parametric regression has several assumptions, while the only one in nonparametric regression is that the shape of the curve does not form a certain pattern. While the study aims to model poverty, the data is per region, which has a fluctuating nature. The data in this research does not form a certain pattern; therefore, nonparametric regression is used. Based on the characteristics of the data, it is very suitable to use the Fourier series approach. In this research, nonparametric regression modeling with one, two, and three oscillation parameters was attempted. The best model was obtained, which consisted of two oscillation parameters with a GCV value of 2.110 and a coefficient of determination of 92.44%.

Keywords: *Fourier Series, GCV, Nonparametric Regression, Oscillation Parameters, Poverty.*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Provinsi Sumatera Barat adalah provinsi yang terus mengupayakan pengentasan kemiskinan. Persentase kemiskinan di Provinsi Sumatera Barat dari tahun 2017 hingga tahun 2020 mengalami penurunan yang terjadi dari tahun ke tahun. Akan tetapi, persentase kemiskinan mengalami peningkatan sebesar 0,35 persen pada Tahun 2021 dibandingkan Tahun 2020 (BPS, 2021). Peningkatan persentase kemiskinan ini menjadikan pemerintah perlu membuat kebijakan untuk menurunkan persentase kemiskinan di tahun berikutnya. Berdasarkan data diketahui bahwa daerah kabupaten memiliki tingkat kemiskinan lebih tinggi daripada daerah perkotaan baik di Tahun 2017 hingga Tahun 2021. Daerah dengan tingkat kemiskinan tertinggi pada Tahun 2021 adalah Kabupaten Kepulauan Mentawai sebesar 14,84% (BPS, 2021). Menurut Leasiwal (2013), kemiskinan perlu menjadi pusat perhatian karena berhubungan dengan kondisi sosial dan kesejahteraan masyarakat. Terdapat banyak hal yang menyebabkan pengentasan masalah kemiskinan sulit untuk dicapai. Hal tersebut perlu dianalisis sehingga pemerintah dapat memaksimalkan kebijakan untuk menangani kemiskinan agar tingkat kemiskinan bisa ditekan. Beberapa faktor yang menyebabkan kemiskinan adalah pendidikan (Seruni, 2014) dan pekerjaan (Eddy, 2010).

Analisis regresi merupakan salah satu metode yang umum digunakan dalam memodelkan faktor-faktor yang mempengaruhi persentase kemiskinan. Pada analisis ini, terdapat beberapa pendekatan untuk mengestimasi kurva regresi seperti parametrik, nonparametrik dan semiparametrik. Regresi parametrik dalam analisisnya mempunyai asumsi baku yang sulit untuk dipenuhi (Budiantara, 2009). Sedangkan, regresi nonparametrik mengasumsikan bentuk kurva regresi mulus (*smooth*). Kurva regresi dari data yang digunakan pada penelitian ini tidak membentuk suatu pola dan cenderung berulang. Apabila kurva tidak mengikuti pola tertentu maka tidak dapat dianalisis menggunakan regresi parametrik. Oleh karena itu, untuk analisisnya digunakan regresi nonparametrik dengan pendekatan Deret Fourier (Eubank, 1999).

Regresi nonparametrik Deret Fourier ialah polinomial yang memuat fungsi trigonometri dan memiliki kemampuan yang baik dalam menangani pola data yang cenderung berulang. Penelitian pertama mengenai regresi nonparametrik Deret Fourier dilakukan pada tahun 1992 oleh Bolideau. Hal terpenting dalam analisis ini yaitu pemilihan parameter penghalus yang optimal sehingga diperoleh hasil yang mendekati keadaan yang sebenarnya (Budiantara, 1994; Hardle, 1990). Tujuan dari penelitian ini yaitu memodelkan dan mengetahui bentuk grafik dari persentase kemiskinan di Provinsi Sumatera Barat menggunakan regresi nonparametrik estimator Deret Fourier. Beberapa penelitian terdahulu mengenai analisis ini antara lain, Prahutama (2013) melakukan penelitian pada kasus tingkat pengangguran terbuka di Jawa Timur dan diperoleh akurasi sebesar 96,67%. Selanjutnya, Nurjanah (2015)

melakukan penelitian pada data curah hujan di Kota Semarang dan diperoleh akurasi sebesar 92%. Kemudian, Octavanny (2021) meneliti pemodelan anak pernah lahir di Indonesia dan diperoleh akurasi sebesar 80,04%.

II. METODE PENELITIAN

A. Analisis Regresi

Analisis regresi ialah suatu cara yang digunakan untuk melihat pengaruh ataupun hubungan antara variabel yang mempengaruhi terhadap variabel yang dipengaruhi. Selain itu, analisis ini digunakan juga untuk memprediksi (Budiantara, 2009). Berdasarkan jumlah variabelnya regresi dibedakan menjadi regresi linier sederhana, regresi linier berganda, regresi multirespon dan regresi multivariabel. Dalam analisis regresi terdapat tiga pendekatan untuk mengestimasi kurva regresi yaitu pendekatan parametrik, nonparametrik dan semiparametrik (Rencher, 2002).

B. Regresi Parametrik

Regresi parametrik merupakan metode yang digunakan untuk mengetahui pola hubungan antara variabel dependen dengan variabel independen dengan asumsi bentuk kurva regresi membentuk pola tertentu. Dalam penggunaannya, analisis ini memiliki asumsi yang sulit untuk dipenuhi. Sehingga, apabila ada asumsi yang tidak terpenuhi maka hasil yang diperoleh tidak terjamin keakuratannya. Oleh karena itu, solusinya digunakan regresi nonparametrik (Budiantara, 2009; Suparti, 2018). Menurut Montgomery (2012), model regresi parametrik disajikan pada Persamaan (1).

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad (1)$$

C. Regresi Nonparametrik

Regresi nonparametrik digunakan ketika memodelkan data yang tidak diketahui bentuk pola dari kurva regresinya. Regresi nonparametrik mengharapkan data mencari sendiri bentuk estimasi dari kurva regresinya tanpa dipengaruhi oleh pandangan peneliti. Regresi nonparametrik memiliki fleksibilitas yang tinggi karena fungsi regresi sebagai representasi kurva regresi dapat menghampiri suatu pola data berpasangan yang secara visual disajikan dalam *scatterplot* (Eubank, 1999). Model regresi nonparametrik disajikan pada Persamaan (2).

$$y_i = \underline{f}(x_i) + \varepsilon_i \quad (2)$$

y_i menjelaskan variabel dependen pada observasi ke- i , x_i menjelaskan variabel independen pada observasi ke- i , dan f fungsi regresi yang polanya tidak dapat ditentukan (Eubank, 1999; Mendenhall, 2012).

D. Regresi Nonparametrik Deret Fourier

Penelitian pertama mengenai regresi nonparametrik estimator Deret Fourier dilakukan pada Tahun 1992 oleh Bilodeau. Analisis ini ialah polinomial trigonometri yang mempunyai kemampuan untuk menyesuaikan diri terhadap sifat asli data. Kelebihan dari analisis ini yaitu dapat menangani pola data yang cenderung berulang, dapat digunakan ketika data menunjukkan pola yang bergelombang dan mampu mengatasi data dengan pola musiman (Bilodeau, 1992; Pane *et.al.*, 2020).

Hubungan antara data variabel independen ($x_{1i}, x_{2i}, \dots, x_{mi}$) dengan variabel dependen y_i secara nonparametrik dijelaskan oleh model pada Persamaan (3),

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{mi}) + \varepsilon_i \quad (3)$$

Pemodelan regresi nonparametrik yang digunakan merupakan kombinasi aditif fungsi linier dan cosinus. Menurut Bilodeau (1992), fungsi f didekati dengan fungsi Deret Fourier yang disajikan pada Persamaan (4).

$$f_j(x_{ji}) = \frac{1}{2}\alpha_0 + b_j x_{ji} + \sum_{k=1}^K \alpha_{kj} \cos_k x_{ji}, \quad j = 1, 2, \dots, m \quad (4)$$

E. Estimasi Parameter Model

Estimasi parameter model dihitung menggunakan metode *Least Square* yaitu dengan meminimumkan jumlah kuadrat *residual* yang diperoleh dari model. Sehingga, model yang dihasilkan mampu menjelaskan data dengan baik. Menurut Eubank (1999), model regresi pada Persamaan (4) disajikan dengan matriks sehingga diperoleh Persamaan (5),

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cos x_{11} & \dots & \cos_k x_{11} & \vdots & \dots & \vdots & x_{j1} & \cos x_{j1} & \dots & \cos_k x_{j1} \\ 1 & x_{12} & \cos x_{12} & \dots & \cos_k x_{12} & \vdots & \dots & \vdots & x_{j2} & \cos x_{j2} & \dots & \cos_k x_{j2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cos x_{1n} & \dots & \cos_k x_{1n} & \vdots & \dots & \vdots & x_{jn} & \cos x_{jn} & \dots & \cos_k x_{jn} \end{bmatrix} \begin{bmatrix} \frac{1}{2}\alpha_0 \\ b_j \\ \vdots \\ \alpha_{kj} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (5)$$

berdasarkan Persamaan (5) maka diperoleh,

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad (6)$$

sehingga didapatkan penduga dari $\underline{\hat{\beta}}$ yang disajikan pada Persamaan (7).

$$\underline{\hat{\beta}} = (\underline{X}'_k \underline{X}_k)^{-1} \underline{X}'_k \underline{y} \quad (7)$$

F. Parameter Osilasi

Parameter osilasi atau parameter penghalus (λ) merupakan pengontrol keseimbangan fungsi f . Nilai λ yang besar akan menyebabkan kurva estimator fungsi f semakin *smooth* atau *overfitting*, keadaan ini menyebabkan kurva regresi yang terbentuk menjauhi pola data sebenarnya. Sebaliknya, apabila nilai λ kecil maka kurva estimator fungsi f menjadi kasar atau *underfitting*, sehingga galat yang diperoleh sangat besar. Nilai λ yang baik adalah nilai λ yang bernilai tidak tinggi dan tidak rendah (Wahba, 1985). Nilai λ yang dicobakan tidak memiliki batasan, akan tetapi semakin banyak nilai λ yang dicobakan, model yang dihasilkan akan semakin kompleks. Salah satu metode yang digunakan dalam menentukan nilai λ yang optimal adalah *Generalized Cross Validation* (GCV) (Eubank, 1999). Rumus metode GCV disajikan pada Persamaan (8).

$$GCV = \frac{MSE(\lambda)}{(n^{-1} \text{trace}(\mathbf{I} - \mathbf{H}(x)))^2} \quad (8)$$

$MSE(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, dengan \mathbf{H} merupakan matriks Hessian, \mathbf{I} merupakan matriks identitas, n menunjukkan jumlah pengamatan, y_i ialah variabel dependen ke- i dan \hat{y}_i ialah prediksi variabel dependen ke- i .

G. Koefisien Determinasi

Koefisien determinasi (R^2) ialah alat yang digunakan untuk mengetahui ketepatan model yang diperoleh sehingga dapat diketahui seberapa besar variabel independen dapat menjelaskan naik atau turunnya variabel dependen. Adapun untuk menghitung koefisien determinasi dapat menggunakan rumus pada Persamaan (9).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

y_i ialah variabel dependen pada observasi ke- i , \hat{y}_i ialah hasil prediksi variabel dependen pada observasi ke- i , dan \bar{y} ialah rata-rata variabel dependen. Nilai koefisien determinasi yang baik adalah nilai koefisien determinasi yang mendekati 1 (Montgomery, 2012).

H. Sumber Data dan Teknik Analisis Data

Data pada penelitian ini adalah data persentase kemiskinan beserta faktor-faktor yang diduga mempengaruhinya yang diperoleh dari publikasi Badan Pusat Statistik Provinsi Sumatera Barat Tahun 2021 yang diakses melalui website resmi. Variabel yang digunakan adalah persentase kemiskinan (Y), rata-rata lama sekolah (X_1), bekerja di sektor pertanian (X_2) dan bekerja di sektor informal (X_3).

Adapun langkah-langkah analisis regresi nonparametrik estimator Deret Fourier adalah sebagai berikut.

1. Menyajikan deskripsi dari masing-masing variabel yang digunakan.
2. Mengetahui pola data dengan membuat *scatterplot* variabel dependen (Y) dengan setiap variabel independen (X).
3. Memodelkan data dengan regresi nonparametrik Deret Fourier menggunakan satu parameter osilasi, dua parameter osilasi dan tiga parameter osilasi.
4. Menghitung nilai GCV dari masing-masing parameter osilasi yang dicobakan.
5. Mendapatkan model terbaik berdasarkan parameter osilasi yang memperoleh nilai GCV minimum.
6. Membuat grafik dari model terbaik yang telah diperoleh.
7. Menghitung nilai R^2 dari model terbaik.

III. HASIL DAN PEMBAHASAN

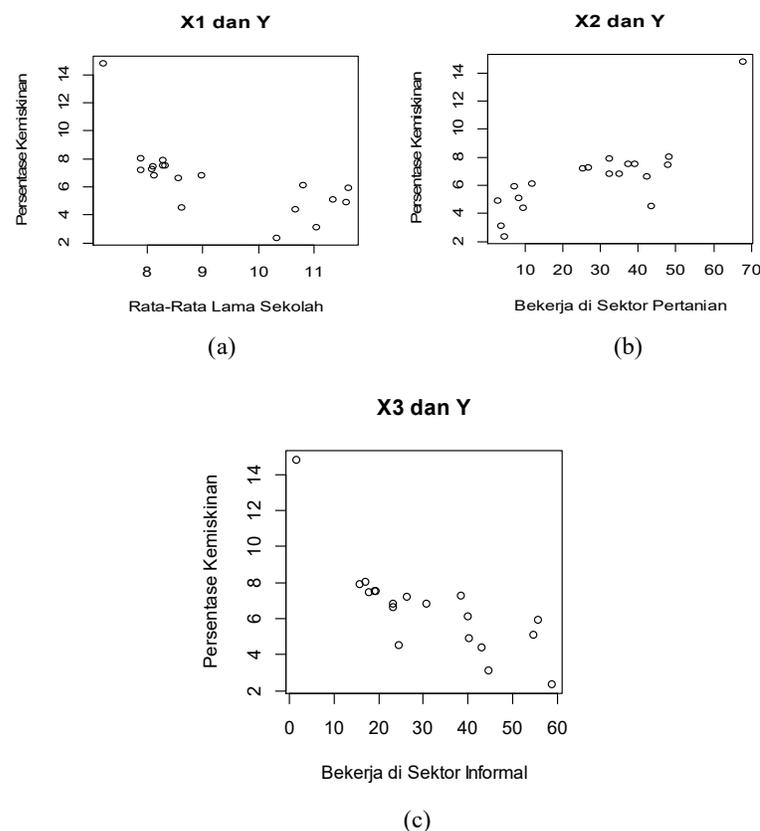
A. Statistika Deskriptif

Langkah pertama sebelum melakukan pemodelan pada data yang digunakan yaitu menyajikan statistika deskriptif dari data untuk mengetahui karakteristik dari setiap variabel yang digunakan. Statistika deskriptif dari setiap variabel disajikan pada Tabel 1.

Tabel 1. Statistika deskriptif variabel penelitian

Variabel	N	Maksimum	Minimum	Rata-Rata	Standar Deviasi
Y	19	14.8	2.38	6.56	2.57
X ₁	19	11.63	7.2	9.24	1.49
X ₂	19	67.85	2.63	27.52	18.84
X ₃	19	58.82	1.56	31.23	15.66

Berdasarkan Tabel 1 diketahui bahwa setiap kabupaten/kota di Provinsi Sumatera Barat memiliki karakter yang berlainan pada setiap variabel independennya. Kabupaten Kepulauan Mentawai adalah daerah dengan tingkat kemiskinan terbesar yaitu sebesar 14,8%. Sedangkan, daerah dengan persentase kemiskinan terkecil yaitu Kota Sawahlunto sebesar 2,38%. Selanjutnya, pola hubungan antara persentase kemiskinan dengan masing-masing variabel independennya disajikan menggunakan *scatterplot* seperti yang ditampilkan pada Gambar 1.



Gambar 1. (a) Persentase kemiskinan dan rata-rata lama sekolah, (b) Persentase kemiskinan dan bekerja di sektor pertanian, (c) Persentase kemiskinan dan bekerja di sektor informal

Berdasarkan Gambar 1 diketahui bahwa data menunjukkan tidak adanya kecenderungan membentuk suatu pola. Oleh karena itu, tidak akurat apabila dianalisis menggunakan regresi parametrik. Solusinya, untuk pola data yang tidak membentuk suatu pola dan cenderung berulang dianalisis menggunakan regresi nonparametrik estimator deret fourier (Eubank, 1999). Penelitian ini menggunakan satu sampai dengan tiga parameter osilasi.

B. Pemodelan Regresi Nonparametrik Estimator Deret Fourier

Tahap awal dalam pemodelan regresi nonparametrik dengan pendekatan deret fourier yaitu menentukan nilai parameter osilasi yang dilihat berdasarkan nilai GCV terkecil. Parameter osilasi dilambangkan sebagai K dan berupa bilangan asli. Nilai K yang digunakan dari satu sampai tiga. Nilai K yang digunakan terbatas pada tiga

parameter osilasi karena dengan menggunakan tiga parameter osilasi hasil koefisien determinasi yang diperoleh sangat tinggi yaitu 95,95%. Adapun output yang diperoleh dari proses pemodelan persentase kemiskinan (Y) disajikan pada Tabel 2.

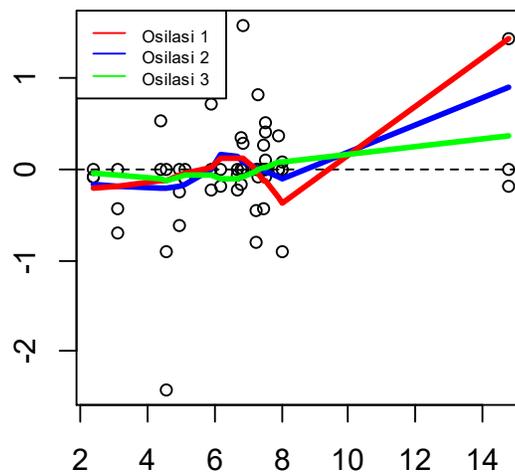
Tabel 2. Nilai GCV masing-masing parameter osilasi

Nilai K	GCV	MSE	Koefisien Determinasi
$K = 1$	8.392	0.836	86.63%
$K = 2$	2.110	0.473	92.44%
$K = 3$	0.636	0.254	95.95%

Berdasarkan Tabel 2, diperoleh bahwa nilai GCV yang paling rendah yaitu 0,636 dengan nilai $K=3$. Akan tetapi, nilai GCV yang terlalu rendah akan menghasilkan model yang sangat mulus. Sebaliknya, nilai GCV tertinggi diperoleh ketika nilai $K= 1$, nilai GCV yang tinggi menyebabkan model yang dihasilkan kasar. Model terbaik adalah model yang tidak menyebabkan terjadinya *overfitting* ataupun *underfitting*. Oleh karena itu, model terbaik yang dipilih adalah model dengan nilai $K=2$. Adapun hasil estimasi parameter menggunakan dua parameter osilasi adalah sebagai berikut.

$$\hat{y} = 19.992 - 0.508x_{1i} + 2.636 \cos x_{1i} + 0.934 \cos 2x_{1i} - 0.084x_{2i} + 0.632 \cos x_{2i} - 1.196 \cos 2x_{2i} - 0.180x_{3i} + 0.445 \cos x_{3i} - 1.037 \cos 2x_{3i}$$

Setelah diperoleh model, sehingga dapat disajikan *scatterplot* yang memuat kurva regresi yang berisi masing-masing nilai K yang dicobakan untuk memastikan model terbaik yang terpilih seperti yang ditampilkan pada Gambar 2.



Gambar 2. Plot perbandingan antara masing-masing nilai K

Berdasarkan Gambar 2 dapat dipastikan bahwa model terbaik adalah model dengan nilai $K=2$. Meskipun nilai GCV terendah dan nilai R^2 tertinggi diperoleh ketika $K=3$. Hal ini dikarenakan, kurva regresi dengan tiga parameter osilasi menyebabkan terjadinya *overfitting* dimana kurva regresinya tidak mengikuti pola data asli. Sehingga, model terbaik yang dipilih adalah model dengan dua parameter osilasi.

IV. KESIMPULAN

Model terbaik yang terpilih adalah model dengan dua parameter osilasi. Grafik model terbaik ditandai dengan warna biru. Berdasarkan grafik tersebut dapat diprediksi penurunan atau peningkatan persentase kemiskinan di tahun berikutnya berdasarkan persebaran data. Ketika persebaran data berada dibawah grafik maka diidentifikasi bahwa persentase kemiskinan mengalami penurunan. Sedangkan, ketika persebaran data berada diatas grafik maka diidentifikasi bahwa persentase kemiskinan mengalami peningkatan. Hal ini berguna untuk mendeteksi dini peningkatan persentase kemiskinan. Sehingga, pemerintah dapat menerapkan kebijakan yang akan dilakukan dalam menangani masalah kemiskinan pada waktu berikutnya. Hasil pemodelan menunjukkan bahwa analisis ini efektif

dalam memodelkan persentase kemiskinan di Provinsi Sumatera Barat. Hal ini ditunjukkan dengan nilai koefisien determinasi yang diperoleh sebesar 92,44% yang artinya variabel rata-rata lama sekolah, bekerja di sektor pertanian dan bekerja di sektor informal mampu menjelaskan keragaman persentase kemiskinan sebesar 92,44%. Penelitian berikutnya disarankan untuk memakai metode lain ataupun melakukan perbandingan antara metode regresi nonparametrik Deret Fourier dengan metode lain.

DAFTAR PUSTAKA

- Badan Pusat Statistik (2021), *Berita Resmi Statistik No. 07/01/Th.XXV, 17 Januari 2022: Profil Kemiskinan di Indonesia September 2021*, Badan Pusat Statistik, Jakarta.
- _____ (2022), *Provinsi Sumatera Barat dalam Angka 2022*, Badan Pusat Statistik Sumatera Barat, Padang.
- Bilodeau, M. (1992). "Fourier Smoother and Additive Models", *The Canadian Journal of Statistics*, Vol. 3, hal. 257-269.
- Budiantara, I. N. (2009). *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Surabaya: Institut Teknologi Sepuluh November.
- Eddy S. S., dan Agung. (2010). "Faktor-Faktor yang Mempengaruhi Kemiskinan secara Makro di Lima Belas Provinsi Tahun 2007", *Jurnal dan Organisasi Manajemen*, Vol. 6, No. 2, hal. 89-100.
- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing, Second Edition*. New York: Marcel Dekker.
- Hardle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.
- Leasiwal, C. T. (2013). "Determinan dan Karakteristik Kemiskinan di Provinsi Maluku", *Jurnal Ekonomi Cita Ekonomika*, Vol. 7, No. 2, hal. 1-26.
- Mendenhall, W., dan Sincich, T. (2012). *A Second Course in Statistics Regression Analysis*. New York: Prentice Hall.
- Montgomery, D. C., Peck, E. A., dan Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Canada: John Wiley & Sons, Inc.
- Nurjanah, F., Utami, T. W., dan Nur, I. M. (2015). "Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Pola Data Curah Hujan di Kota Semarang", *Jurnal Statistika Universitas Muhammadiyah Semarang*, Vol. 3, No. 2, hal. 8-14.
- Octavanny, M. A. D., Budiantara, I. N., Kuswanto, H., dan Rahmawati, D. P. (2021). "Modelling Children Ever Born in Indonesia Using Fourier Series Nonparametric Regression", *Journal of Physics: Conference Series*, Vol. 1752, No. 1, hal. 1-7.
- Pane, R. S., dan Ampa, A. T. (2020). "Estimation of Heteroskedasticity Semiparametric Regression Curve Using Fourier Series Approach", *Journal of Research in Mathematics Trends and Technology (JoRMTT)*, Vol. 2, No. 1, hal. 14-20.
- Prahatama, 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.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis, Second Edition*. New York: John Wiley & Sons, Inc.
- Seruni P. S. P., dan Sutrisna, K. (2014). "Pengaruh PDRB per Kapita, Pendidikan dan Produktivitas Tenaga Kerja terhadap Kemiskinan di Provinsi Bali", *E-Jurnal EP Unud*, Vol. 3, No. 10, hal. 431-439.
- Suparti, Santoso, R., Prahatama, A., dan Devi, A. R. (2018). *Regresi Nonparametrik*. Jawa Timur: Wade Group National Publishing.
- Wahba, G. (1985). "A Comparison of GCV and GML for Choosing the Smoothing Parameter in Generalized Spline Smoothing Problem", *The Annal of Statistic*, Vol. 13, No. 4, hal. 1378-1402.
- _____ (1990). *Spline Models For Observation Data*. Philadelphia: SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics.