https://journal.uinsgd.ac.id/index.php/kubik/index
Vol. 10, No. 2, 2025, pp. 141-151
e-ISSN: 2686-0341 p-ISSN: 2338-0896

Modeling The Open Unemployment Rate Using A Ridge-Fourier Series Estimator In Nonparametric Regression

Sidratul Muthahharah[1)], I Nyoman Budiantara[1) *)], Vita Ratnasari[1)]

[1]*Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[*)]*email:* *i_nyoman_b@statistika.its.ac.id*

Abstract

Nonparametric regression is widely used to model relationships between variables when the functional form of the regression curve remains unspecified. The Fourier series estimator is effective for capturing periodic and complex nonlinear patterns. However, in multivariable settings, strong correlations among predictors may lead to multicollinearity, resulting in unstable parameter estimates due to the near-singularity of the design matrix. Although Fourier-based estimators have been extensively developed, their formulation has not explicitly addressed this issue. This study proposes a ridge-Fourier series estimator for nonparametric regression to obtain stable parameter estimation in the presence of multicollinearity. The estimator is derived within a penalized likelihood framework by incorporating a ridge penalty into the Fourier series representation and estimating the parameters. The oscillation and ridge penalty parameters are selected simultaneously using the Generalized Cross Validation (GCV) criterion. The proposed method is applied to the 2024 Open Unemployment Rate data covering districts and cities in West Java. The optimal model is obtained at the oscillation combination of [4, 4, 4, 2] with a ridge parameter of 0.0336, accompanied by a coefficient determination of 91.1%, indicating that the proposed ridge-Fourier estimator provides more stable estimation and improved predictive performance under multicollinearity conditions.

Keywords: Nonparametric Regression, Fourier Series, Ridge Regression, Multicollinearity, GCV
MSC2020: 62G05, 62P20, 62P25

Abstrak

Regresi nonparametrik banyak digunakan untuk memodelkan hubungan antar variabel ketika bentuk fungsi regresi tidak diketahui pasti. Estimator deret Fourier efektif dalam menangkap pola periodik dan nonlinear yang kompleks. Namun, pada kasus multivariabel, korelasi yang kuat antar variabel prediktor dapat menyebabkan multikolinearitas yang mengakibatkan estimasi parameter menjadi tidak stabil akibat matriks desain yang mendekati singular. Meskipun estimator berbasis Fourier telah banyak dikembangkan, formulasinya belum secara eksplisit mengatasi permasalahan tersebut. Penelitian ini mengusulkan estimator ridge–deret Fourier dalam regresi nonparametrik untuk memperoleh estimasi parameter yang stabil pada kondisi multikolinearitas. Estimator diturunkan dalam kerangka likelihood terpinalti dengan mengintegrasikan penalti ridge ke dalam representasi deret Fourier. Parameter osilasi dan penalti ridge dipilih secara simultan menggunakan kriteria Generalized Cross Validation (GCV). Metode yang diusulkan diterapkan pada data Tingkat Pengangguran Terbuka tahun 2024 pada kabupaten/kota di Jawa Barat. Model optimal diperoleh pada kombinasi osilasi [4, 4, 4, 2] dengan parameter ridge sebesar 0,0336 dan koefisien determinasi sebesar 91,1%, *menunjukkan bahwa estimator ridge–Fourier yang diusulkan mampu memberikan*

*) Corresponding Author
Received: 20-03-2026, Accepted: 20-04-2026, Published: 27-04-2026

estimasi parameter yang lebih stabil serta meningkatkan kinerja prediksi pada kondisi multikolinearitas.

Keywords: Nonparametric Regression, Fourier Series, Ridge Regression, Multicollinearity, GCV
MSC2020: 62G05, 62P20, 62P25

Citation: S. Muthahharah, I. N. Budiantara, and V. Ratnasari, "*Modeling The Open Unemployment Rate Using A Ridge-Fourier Series Estimator in Nonparametric Regression*", *KUBIK J. Publ. Ilm. Mat.*, Vol. 9, No. 1, pp. 1-9, 2024.

Introduction

Regression is a statistical method commonly used to analyze the influence of predictor variables on response variables, as well as to investigate the estimated form of the regression curve. Several approaches are frequently employed to explore the estimated form of a regression curve, one of which is the nonparametric approach. Nonparametric regression serves as an alternative for investigating the form of a regression curve when the data pattern cannot be identified using a scatter plot [1]. Various estimators for nonparametric regression curves have been developed, including spline, Fourier series, and kernel estimators, among others. Among the estimators developed for regression model estimation, the Fourier series has become one of the most widely used due to its ability to handle cyclical data patterns using trigonometric functions, particularly the cosine function [2].

In classical regression methods, one of the main assumptions that must be satisfied is the absence of multicollinearity, as its presence can lead to invalid regression analysis [3]. In practice, however, it is common to encounter situations where predictor variables are highly correlated with one another, resulting in large variances in parameter estimation [4]. According to Montgomery [5], several methods can be employed to address multicollinearity, including additional data collection, model respecification, ridge regression, and principal component regression. Among these methods, ridge regression is the most widely recommended approach. Ridge regression is a method designed to obtain stable regression coefficient estimates in the presence of multicollinearity. It works by handling the strong relationships among independent variables in the regression model, which can cause the coefficient matrix to approach singularity and consequently lead to instability in parameter estimation [6].

Research on nonparametric regression using the Fourier series function has been extensively carried out, as demonstrated in several previous studies [2],[7],[8],[9],[10]. However, the handling of multicollinearity in nonparametric regression models has not been thoroughly addressed, particularly when the model involves factors that are potentially highly correlated with one another. Ridge regression is a preferred approach for dealing with multicollinearity due to its ability to stabilize estimation, maintaining a balance between bias and variance, as conducted in previous studies [11],[12],[13]. It is noteworthy that the existing Fourier series-based studies [2],[7],[8],[9] focus exclusively on estimator construction and model fit, without addressing the issue of multicollinearity among predictor variables. In these formulations, the design matrix $X(S)$ retains the same collinearity structure as in classical regression; when predictor variables are strongly correlated, the matrix $X'(S)X(S)$ becomes nearly singular and the resulting parameter estimates are unstable, regardless of whether a Fourier basis is used.

Unlike existing Fourier series estimators [2],[7],[8],[9], which do not account for multicollinearity among predictor variables, and unlike ridge regression developed in parametric settings [11],[12],[13], the proposed ridge–Fourier estimator addresses both challenges within a single nonparametric framework capturing periodic and nonlinear patterns while stabilizing parameter estimates under strong predictor correlation. This represents the primary methodological contribution of the present study. Drawing on the stabilizing properties of ridge regression [6],[11],[13], this research develops a ridge-Fourier series estimator capable of addressing multicollinearity in data patterns that exhibit periodic and cyclical behavior. To demonstrate the practicality of the proposed method, it is applied to the 2024 Open Unemployment Rate data across districts and cities in West Java Province, Indonesia, where socioeconomic predictor variables exhibit strong multicollinearity and complex nonlinear relationships. Accordingly, this study aims to derive a ridge-Fourier series estimator for nonparametric regression using the Maximum Likelihood Estimation (MLE) approach, with the GCV criterion employed to simultaneously determine the optimal oscillation and ridge parameters.

Methods

Data and Research Variables

This study utilizes secondary data sourced from publications of the Badan Pusat Statistik (BPS) of West Java. The dataset consists of the Open Unemployment Rate for 2024 spanning 27 districts and cities in West Java. The observation units comprise 27 districts/cities within West Java. The study employs one response variable, namely the Open Unemployment Rate, and four predictor variables consisting of Regional Expenditure, Gross Regional Domestic Product, Population Percentage, and Human Development Index.

Table 1. Research Variables

Variable	Description	Type
y	Open Unemployment Rate	Response
x_1	Regional Expenditure	Predictor
x_2	Gross Regional Domestic Product	Predictor
x_3	Population Percentage	Predictor
x_4	Human Development Index	Predictor

Theoretical Framework for Methodology

This subsection explains the theoretical foundation used in this study to model the Open Unemployment Rate using a nonparametric regression approach grounded in Fourier series, incorporating ridge regularization and optimal parameter selection through Generalized Cross Validation (GCV).

Nonparametric Regression

Nonparametric regression is applied to characterize the relationship between a dependent variable and independent variables without assuming a specific functional form. This approach is suitable for modeling data that exhibit nonlinear and complex patterns [10]. The general nonparametric regression model is defined as [14].

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

Fourier Series Regression

The unknown regression function $g(x_i)$ can be approximated using a Fourier series, particularly when the data exhibit periodic or cyclical behavior [7]. The Fourier series approximation is given by [2].

$$g(x_i) = bx_i + \frac{1}{2}a_0 + \sum_{s=1}^{S} a_s \cos sx_i \ , s = 1,2, \dots , S \tag{2}$$

The oscillation parameter S controls the number of cosine components included in the model, where a larger S allows the model to capture more complex and rapidly fluctuating patterns in the data. Thus, the regression model becomes,

$$y_i = bx_i + \frac{1}{2}a_0 + \sum_{s=1}^{S} a_s \cos sx_i + \varepsilon_i \, , s = 1,2, \dots , S \tag{3}$$

For the multivariable case,

$$y_i = \sum_{l=1}^{q} \left(b_l x_{li} + \frac{1}{2}a_{0l} + \sum_{s=1}^{S} a_{sl} \cos sx_{li} \right) + \varepsilon_i \tag{4}$$

In matrix form, the model can be written as,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{5}$$

with the error term,

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\beta} \tag{6}$$

The least squares estimator is obtained as,

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X^T X}\right)^{-\mathbf{1}} \boldsymbol{X^T y} \tag{7}$$

Ridge Regression

When multicollinearity occurs among predictor variables, the matrix $(\boldsymbol{X^T X})$ becomes nearly singular, resulting in unstable parameter estimates [15]. To address this issue, ridge regression introduces a penalty term to the estimation process [6]. The L2 penalty form $\lambda||\beta||^2$ is chosen in this study because it produces a closed-form solution and ensures the augmented matrix $X'(S)X(S) + \lambda I$ remains invertible, thereby stabilizing the coefficient estimates under multicollinearity. Furthermore, unlike the L1 penalty which may eliminate predictor variables, the L2 penalty retains all predictor variables in the model. The objective function is defined as,

$$min\{Q(\boldsymbol{\beta})\} = min\left\{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\varepsilon}_{\boldsymbol{i}}^{\boldsymbol{2}} + \lambda||\boldsymbol{\beta}||^2\right\} = min\{(\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{X\beta}) + \boldsymbol{\lambda\beta^T\beta}\} \tag{8}$$

By minimizing this function, the ridge estimator is obtained as,

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{ridge}} = \left(\boldsymbol{X^T X} + \lambda \boldsymbol{I}\right)^{-\mathbf{1}} \boldsymbol{X^T y} \tag{9}$$

where λ is the ridge parameter controlling the level of regularization [16].

Ridge-Fourier Nonparametric Model

Combining Fourier series with ridge regression produces a more robust model capable of capturing nonlinear patterns while handling multicollinearity. The model is expressed as,

$$y_i = a_0^* + \sum_{l=1}^{q} \left(b_l x_{li} + \sum_{s=1}^{S} a_{sl} \cos sx_{li} \right) + \varepsilon_i \, ; i = 1,2, \dots , n. \tag{10}$$

Equation (10) is obtained by simplifying the multivariable Fourier regression model in Equation (4). In this formulation, the intercept term is redefined as, $a_0^* = \sum_{l=1}^{q} \frac{1}{2} a_{0l}$.

Parameter estimation is carried out within a penalized likelihood framework, where the ridge penalty term λ||β||² is added to the log-likelihood function to penalize large coefficient values and stabilize estimation under multicollinearity. The error terms ε_i are assumed to be independently and identically distributed with $\varepsilon_i \sim N(0, \sigma^2)$, where σ^2 is the error variance. Under this assumption, the log-likelihood function can be expressed explicitly as shown in Equation (11). Consequently, maximizing the penalized log-likelihood with respect to β yields the closed-form estimator in Equation (13). The resulting penalized likelihood function is given by,

$$\arg\max_{\boldsymbol{\beta}}\big(\ell(\boldsymbol{\beta}, \lambda)\big) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}(y - \boldsymbol{X}(\boldsymbol{S})\boldsymbol{\beta})^T(y - \boldsymbol{X}(\boldsymbol{S})\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|^2 \tag{11}$$

To obtain the estimator β that maximizes Equation (11), the penalized log-likelihood is differentiated with respect to β and the result is set equal to zero.

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 \tag{12}$$

Solving Equation (12) yields the ridge-Fourier series estimator as follows,

$$\widehat{\boldsymbol{\beta}}(S, \lambda) = (\boldsymbol{X}'(S)\boldsymbol{X}(S) + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}'(S)\boldsymbol{Y} \tag{13}$$

The estimator in Equation (13) represents the ridge-Fourier series estimator for the nonparametric regression model. The addition of the ridge penalty $\lambda\mathbf{I}$ to the matrix $\mathbf{X}'(S)\mathbf{X}(S)$ ensures that the resulting matrix is invertible even under multicollinearity conditions, thereby producing stable and reliable parameter estimates. When $\lambda = 0$, the estimator reduces to the ordinary least squares estimator in Equation (7).

Generalized Cross Validation (GCV)

The selection of optimal parameters, including the oscillation parameter S and ridge parameter λ, is performed using Generalized Cross Validation (GCV) [17]. GCV is preferred in this study because it does not require a fixed specification of the degrees of freedom, which is advantageous for nonparametric models where the effective degrees of freedom vary with both S and λ. The Mean Square Error (MSE) is defined as [18],

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{14}$$

In matrix form,

$$MSE(\lambda, S) = \frac{1}{n}\boldsymbol{y}^{\boldsymbol{T}}\big(\boldsymbol{I} - \boldsymbol{H}(\lambda, \boldsymbol{S})\big)^{\boldsymbol{T}}\big(\boldsymbol{I} - \boldsymbol{H}(\lambda, \boldsymbol{S})\big)\boldsymbol{y} \tag{15}$$

where,

$$\boldsymbol{H}(\lambda, \boldsymbol{S}) = \boldsymbol{X}(\boldsymbol{S})(\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{S})\boldsymbol{X}(\boldsymbol{S}) + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{S})$$

The GCV function is given by,

$$GCV(\lambda, S) = \frac{MSE(\lambda, S)}{\Big(n^{-1}tr\big(\boldsymbol{I} - \boldsymbol{H}(\lambda, \boldsymbol{S})\big)\Big)^2} \tag{16}$$

The optimal model is obtained by,

$$GCV(\lambda, S) = \min_{S}\left\{\frac{MSE(\lambda, S)}{\Big(n^{-1}tr\big(\boldsymbol{I} - \boldsymbol{H}(\lambda, \boldsymbol{S})\big)\Big)^2}\right\} \tag{17}$$

Analysis Steps

This study follows a structured analytical procedure to model the Open Unemployment Rate, identify the relationship between the response and predictor variables, and evaluate the performance of the Ridge-Fourier nonparametric regression model. The analysis is conducted through the following steps,

1. Collect the Open Unemployment Rate data and relevant predictor variables for regencies/cities in West Java, then perform descriptive statistical analysis.
2. Construct scatter plots between the response variable and each predictor variable to examine the pattern of relationships and detect potential nonlinearity.
3. Detect multicollinearity among predictor variables.
4. Model the Open Unemployment Rate using the Ridge-Fourier nonparametric regression approach.
5. Determine the optimal oscillation parameter S and ridge parameter λ using the Generalized Cross Validation (GCV) criterion.
6. Select the best model based on the minimum GCV value.
7. Evaluate the performance of the selected model using the coefficient of determination R^2 and Mean Squared Error (MSE).
8. Draw conclusions and provide recommendations based on the analysis results.

Results and Discussion

Modeling of Open Unemployment Rate using Ridge-Fourier

In this study, the Open Unemployment Rate in regencies/cities of West Java in 2024 is modeled using a Ridge-Fourier nonparametric regression approach. This method is applied to capture nonlinear and fluctuating patterns in the data while addressing multicollinearity among predictor variables. The analysis begins with descriptive statistics to understand the characteristics of the data.

Table 2. Descriptive Statistics

Variable	Minimum	Maximum	Mean	Standard Deviation
Open Unemployment Rate	1.58	8.97	6.52	1.71
Regional Expenditure	7.49	102.94	37.20	22.27
Gross Regional Domestic Product	3.88	293.66	65.79	73.58
Population Percentage	0.42	11.29	3.70	2.44
Human Development Index	68.89	83.75	74.68	4.33

Based on Table 2, the Open Unemployment Rate has a mean of 6.52 and a standard deviation of 1.71, indicating moderate variation across regencies/cities. Gross Regional Domestic Product and Regional Expenditure exhibit relatively high variability, Population Percentage shows moderate variation, while the Human Development Index is more homogeneous across regions.

To examine the relationship patterns between the response and each predictor variable, scatter plots are presented in Figure 1. This visualization is used to identify linear or nonlinear tendencies as a basis for selecting an appropriate modeling approach.

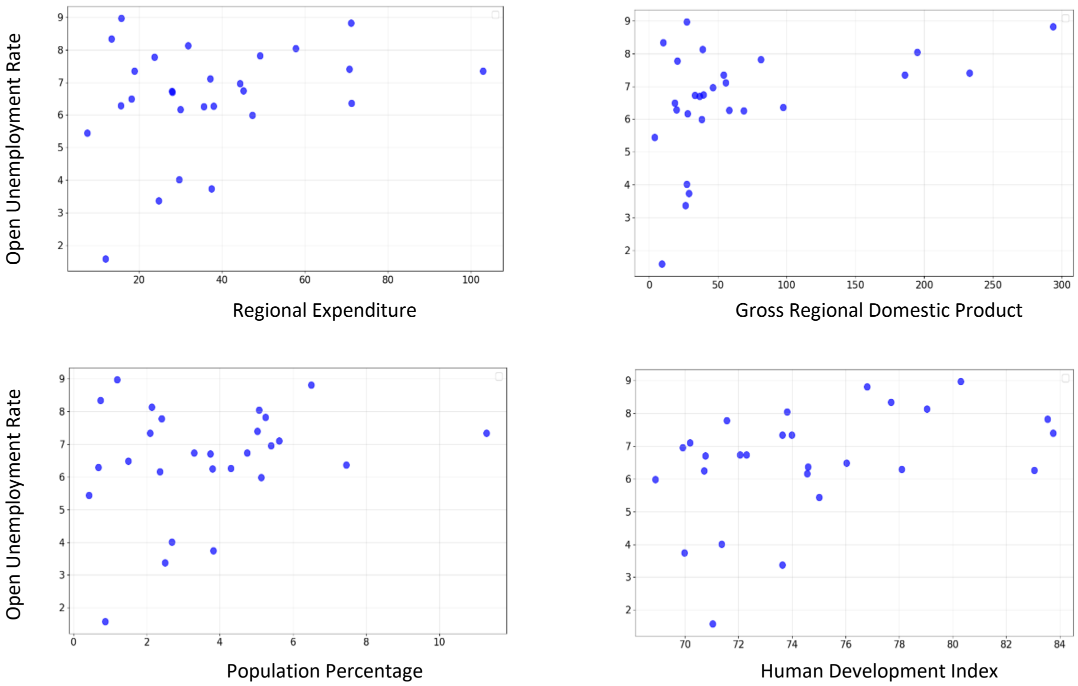

Figure 1. Scatter Plot of Open Unemployment Rate Against Predictor Variables

In particular, the scatter plots of Regional Expenditure (X_1) and Gross Regional Domestic Product (X_2) against the Open Unemployment Rate show more pronounced fluctuating patterns, suggesting stronger nonlinearity compared to the other predictors. This suggests that a flexible modeling approach, such as the Fourier-based nonparametric regression, is appropriate for capturing the underlying structure of the data.

Furthermore, multicollinearity among predictor variables was evaluated using the Variance Inflation Factor (VIF), as presented in Table 3.

Table 3. Variance Inflation Factor (VIF) of Predictor Variables

	X_1	X_2	X_3	X_4
VIF	25.44	4.12	17.43	1.43

Based on Table 3, Regional Expenditure (X_1) and Population Percentage (X_3) have VIF values greater than 10, indicating severe multicollinearity. This condition suggests that the matrix $X'(S)X(S)$ is close to singular, which may result in unstable coefficient estimates under the OLS-based Fourier approach. Therefore, a ridge penalty λI is incorporated to ensure that the augmented matrix remains invertible and yields stable parameter estimates. Based on these findings, the Open Unemployment Rate is subsequently modeled using the ridge-Fourier nonparametric regression framework, as given in Equation (18).

$$\begin{aligned} y_i \;&= a_0^* + b_1 x_{1i} + a_{11}\cos x_{1i} + a_{21}\cos 2x_{1i} + \cdots + a_{S1}\cos Sx_{1i} + \\ &\quad b_2 x_{2i} + a_{12}\cos x_{2i} + a_{22}\cos 2x_{2i} + \cdots + a_{S2}\cos Sx_{2i} + \\ &\quad b_3 x_{3i} + a_{13}\cos x_{3i} + a_{23}\cos 2x_{3i} + \cdots + a_{S3}\cos Sx_{3i} + \\ &\quad b_4 x_{4i} + a_{14}\cos x_{4i} + a_{24}\cos 2x_{4i} + \cdots + a_{S4}\cos Sx_{4i} + \varepsilon_i \end{aligned} \qquad i = 1,2,\dots,n \quad (18)$$

Selection of Optimal Parameters

The optimal values of the Fourier order $S$and ridge penalty λwere determined using the minimum Generalized Cross Validation (GCV) criterion. Various combinations of $S$and λwere evaluated to identify the best-performing model.

Table 4. Minimum GCV, Oscillation Combinations, and Optimal Ridge Parameter

Oscillation Combinations				λ Interval	Minimum GCV	Optimal λ
X_1	X_2	X_3	X_4			
1	1	1	1	[0.0001-0.1]	0.141711	0.1
1	1	1	1	[0.0001-0.1]	0.141954	0.069519
1	1	1	1	[0.0001-0.1]	0.142124	0.048329
1	1	1	1	[0.0001-0.1]	0.142327	0.023357
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	1	1	[0.0001-0.1]	0.15798	0.1
2	2	1	1	[0.0001-0.1]	0.16564	0.1
2	2	2	1	[0.0001-0.1]	0.1614	0.1
2	2	2	2	[0.0001-0.1]	0.16971	0.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	1	1	1	[0.0001-0.1]	0.171454	0.1
3	2	1	3	[0.0001-0.1]	0.212836	0.1
3	3	1	1	[0.0001-0.1]	0.17671	0.1
3	3	3	2	[0.0001-0.1]	0.23159	0.1
3	3	3	3	[0.0001-0.1]	0.256869	0.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
4	4	3	3	[0.0001-0.1]	0.339029	0.1
4	4	4	1	[0.0001-0.1]	0.288225	0.1
4	**4**	**4**	**2**	**[0.0001-0.1]**	**0.100626**	**0.033598**
4	4	4	3	[0.0001-0.1]	0.11362	0.048329
4	4	4	4	[0.0001-0.1]	0.141016	0.069519

Based on Table 4, the optimal model is achieved at the Fourier order combination $[4, 4, 4, 2]$ for the four predictor variables, with the optimal ridge penalty $\lambda = 0.0336$, producing the minimum GCV value of 0.100626. This result suggests that higher-order Fourier components are needed to capture the nonlinear and fluctuating patterns of the Open Unemployment Rate across regencies/cities. The selected parameter combination was then used to estimate the final Ridge-Fourier nonparametric regression model.

Estimation Results of the Ridge-Fourier Model

Using the optimal Fourier order combination $[4, 4, 4, 2]$ and ridge penalty $\lambda = 0.0336$, the estimated Ridge-Fourier nonparametric regression model is obtained as follows,

$$\hat{y}_i = -12.6452 - 0.3868x_{1i} - 1.5503\cos x_{1i} - 0.4267\cos 2x_{1i} - 0.547\cos 3x_{1i} - 1.9885\cos 4x_{1i} + 0.0781x_{2i} - 0.8129\cos x_{2i} - 1.3874\cos 2x_{2i} + 0.6357\cos 3x_{2i} - 0.6166\cos 4x_{2i} + 2.1753x_{3i} - 2.5064\cos x_{3i} - 3.4549\cos 2x_{3i} - 0.3366\cos 3x_{3i} - 2.9666\cos 4x_{3i} +$$

$$0.2690x_{4i} + 1.0692\cos x_{4i} + 3.0676\cos 2x_{4i}$$

The estimated coefficients represent both linear and nonlinear effects of the predictor variables. However, due to the presence of Fourier components, the coefficients are not directly interpretable individually but collectively describe the pattern of the data. This result indicates that the relationship between the Open Unemployment Rate and the predictor variables tends to follow a complex nonlinear pattern.

Interpretation of Results

Both the Fourier order combination [4,4,4,2] and the ridge parameter λ = 0.0336 were determined objectively based on the minimum GCV value, confirming that the selected model is data-driven rather than arbitrarily specified. The results indicate that the Open Unemployment Rate across the 27 regencies/cities of West Java follows complex nonlinear patterns with respect to the four socioeconomic predictors.

The optimal Fourier order combination $[4, 4, 4, 2]$ suggests that Regional Expenditure, Gross Regional Domestic Product, and Population Percentage exhibit more fluctuating relationships with unemployment, while the Human Development Index shows a relatively smoother pattern. This confirms that the predictor–response relationships cannot be adequately represented by simple linear or low-order nonlinear forms.

In addition, the ridge penalty plays an important role in stabilizing estimation under severe multicollinearity, particularly in Regional Expenditure and Population Percentage. The optimal ridge parameter $\lambda = 0.0336$ ensures that the augmented matrix remains invertible and produces reliable coefficient estimates. This relatively small value of λ indicates that only a modest degree of shrinkage was needed to stabilize the estimates, reflecting that multicollinearity in this dataset, while present, does not require aggressive penalization to produce reliable results. From a practical perspective, the high R^2 value of 91.1% indicates that the selected socioeconomic variables explain most of the regional variation in the Open Unemployment Rate, highlighting their importance for labor market policy and regional planning in West Java.

Model Performance Evaluation

The performance of the Ridge-Fourier model was evaluated using the coefficient of determination (R^2) and Mean Squared Error (MSE). Although no train–test split was performed, this study utilizes a complete population dataset of all 27 regencies and cities in West Java, making data partitioning inappropriate. Furthermore, the GCV criterion approximates leave-one-out cross-validation, providing internal validation suitable for small datasets, while the ridge penalty λ directly controls model complexity to reduce the risk of overfitting.

To assess the robustness of the selected ridge parameter, a sensitivity analysis was conducted by varying λ across the range [0.0001, 0.1] while keeping the optimal Fourier order combination [4, 4, 4, 2] fixed. The results show that the GCV value reaches its minimum at λ = 0.0336 and increases gradually as λ deviates from this value in either direction. Similarly, the R^2 value remains relatively stable across the entire range, from 90.0% at λ = 0.1 to 91.3% at very small λ values, with a difference of approximately 1.3 percentage points. These findings confirm that the model is not overly sensitive to the choice of λ and that the selected value produces consistently stable estimates.

The model achieved an R^2 value of 91.1% and an MSE of 0.2497, indicating that the selected predictor variables explain a substantial proportion of the variability in the Open Unemployment Rate. In addition, the MAE value of 0.4170 indicates that the average prediction error is approximately 6.4%

of the mean response value of 6.52%, suggesting that the model produces accurate and reliable predictions across the 27 regencies and cities.

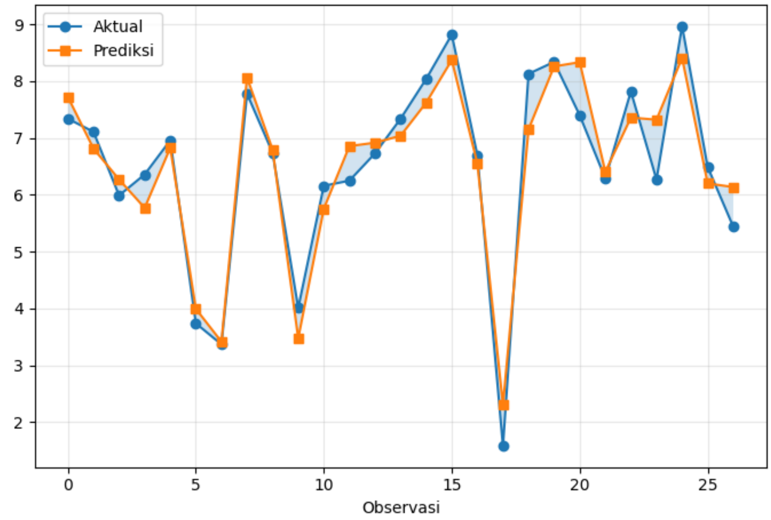

Figure 2. Actual vs. Estimated Open Unemployment Rate Using the Ridge-Fourier Model

In addition, the estimated values are generally close to the observed values, as illustrated in Figure 2. This confirms that the Ridge-Fourier model provides accurate predictions with relatively small errors across most regencies/cities.

Conclusion

This study proposes a Ridge-Fourier series estimator for nonparametric regression to address multicollinearity in data with nonlinear and fluctuating patterns. Applied to the 2024 Open Unemployment Rate across 27 regencies/cities in West Java, the results show that the proposed approach effectively captures complex predictor–response relationships that cannot be adequately represented by conventional linear models.

The optimal model was obtained at the Fourier order combination [4, 4, 4, 2] with a ridge penalty of λ = 0.0336, producing the minimum GCV value and an R^2 of 91.1%. These findings confirm that the Ridge-Fourier estimator not only stabilizes parameter estimation under multicollinearity but also provides strong predictive performance. Therefore, the proposed model offers a flexible and reliable framework for analyzing regional unemployment disparities and can serve as a useful reference for evidence-based policy planning in West Java.

References

[1] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, Second edi. New York: Marcel Dekker, 1999.

[2] M. Bilodeau, "Fourier smoother and additive models," vol. 20, no. 3, pp. 257–269, 1992.

[3] A. Andini, E. Sunandi, P. Novianti, I. Sriliana, and W. Agwil, "Perbandingan Metode Regresi Ridge dan Jackknife Ridge Regression pada Data Tingkat Pengangguran Terbuka," *Limits J. Math. Its Appl.*, vol. 22, no. 1, pp. 77–84, 2025, doi: 10.12962/limits.v22i1.3374.

[4] C. Nisa and S. H. Hastuti, "Kajian Simulasi Perbandingan Metode Ridge Regression Dan Adjusted Ridge Regression Untuk Penanganan Multikolinearitas," *J. Gaussian*, vol. 12, no. 3, pp. 330–339, 2024, doi: 10.14710/j.gauss.12.3.330-339.

[5] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Linear Regression Analysis*, 5th ed. 2012.

[6] W. N. van Wieringen, “Lecture notes on ridge regression,” 2023, [Online]. Available: http://arxiv.org/abs/1509.09169

[7] I. W. Sudiarsa, N. P. A. M. Mariati, N. M. S. Sanjiwani, and D. P. Pramesuari, “Estimator and Applied Mixed Kernel and Fourier Series Modelling in Nonparametric Regression,” *J. Ilm. Ilmu Terap. Univ. Jambi*, vol. 8, no. 2, pp. 622–633, 2024, doi: 10.22437/jiituj.v8i2.36246.

[8] M. I. Salim, Adnan Sauddin, and M. Ichsan Nawawi, “Model Regresi Nonparametrik Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka Di Sulawesi Selatan,” *J. MSA (Mat. dan Stat. serta Apl.*, vol. 10, no. 2, pp. 48–56, 2022, doi: 10.24252/msa.v10i2.30993.

[9] Faisol, S. Hasananh, M. F. F. M., I. Yudistira, and T. Yulianto, “Prediction of seawater salinity based on comparison of truncated spline estimators , Fourier Series and Kernel,” vol. 9, no. 1, pp. 60–68, 2023.

[10] M. Zulfadhli, I. N. Budiantara, and V. Ratnasari, “Nonparametric regression estimator of multivariable Fourier Series for categorical data,” *MethodsX*, vol. 13, no. July, p. 102983, 2024, doi: 10.1016/j.mex.2024.102983.

[11] E.-J. Le and J.-H. Jhong, “Ridge regression splines with knot selection via ℓ_0 penalty,” *J. Korean Data Inf. Sci. Soc.*, vol. 35, no. 6, pp. 889–904, 2024, doi: 10.7465/jkdi.2024.35.6.889.

[12] F. Rahmawati and R. Y. Suratman, “Performa Regresi Ridge dan Regresi Lasso pada Data dengan Multikolinearitas,” *Leibniz J. Mat.*, vol. 2, no. 2, pp. 1–10, 2022, doi: 10.59632/leibniz.v2i2.176.

[13] H. M. Nayem, S. Aziz, and B. M. G. Kibria, “Evaluating Estimator Performance Under Multicollinearity: A Trade-Off Between MSE and Accuracy in Logistic, Lasso, Elastic Net, and Ridge Regression with Varying Penalty Parameters,” *Stats*, vol. 8, no. 2, p. 45, 2025, doi: 10.3390/stats8020045.

[14] M. Ramli, I. N. Budiantara, and V. Ratnasari, “A method for parameter hypothesis testing in nonparametric regression with Fourier series approach,” *MethodsX*, vol. 11, no. October 2023, p. 102468, 2023, doi: 10.1016/j.mex.2023.102468.

[15] N. Khalid, D. M. Khan, M. Suhail, U. Khalil, and E. H. Alkhammash, “Exploring new estimators in ridge regression: Addressing multicollinearity in economic and petroleum product data analysis,” *Kuwait J. Sci.*, vol. 52, no. 4, 2025, doi: 10.1016/j.kjs.2025.100448.

[16] E. Sulistianingsih, Suparti, and D. Ispriyanti, “Pemodelan indeks pembangunan manusia di jawa tengah menggunakan metode regresi ridge dan regresi stepwise,” vol. 11, pp. 468–477, 2022, doi: 10.14710/J.GAUSS.11.3.468-477.

[17] M. N. Amrullah and S. H. Amalia, “COMPARISON OF GENERALIZED CROSS VALIDATION (GCV) METHODS WITH CROSS VALIDATION (CV) TO DETERMINE OPTIMAL KNOTS IN FOURIER SERIES NONPARAMETRIC REGRESSION (Case Study: Poverty Rate in North Sumatra Province),” vol. 3, no. 1, pp. 1–10, 2022, doi: 10.51402/jle.v3i1.5.

[18] S. W. Izumi and Teti Sofia Yanti, “Pemodelan Tingkat Pengangguran Terbuka (TPT) di Jawa Barat Menggunakan Regresi Nonparametrik Deret Fourier,” *Bandung Conf. Ser. Stat.*, vol. 3, no. 2, pp. 594–601, 2023, doi: 10.29313/bcss.v3i2.8769.