Implementation of a MATLAB-Based Computational Framework for Multivariate Linear Regression: A Case Study on Economic Growth in Jayapura

Bobi Frans Kuddi(1), Agung Dwi Saputro(2), Feby Seru(3), Cecilia Bintang Girik Allo(4)

[1,4]Department of Statistics, Universitas Cenderawasih
[2]Department of Information Systems, Universitas Cenderawasih
[3]Department of Mathematics, Universitas Cenderawasih

Kampus Uncen Waena, Jl. Camp Wolker, Kota Jayapura, Papua
e-mail: kuddi198866@gmail.com(1), dwisaputro321@gmail.com(2), febyseru.math@gmail.com(3), caecilia.bintang@fmipa.uncen.ac.id(4)

ABSTRAK

Pertumbuhan ekonomi regional merupakan indikator penting dalam mengevaluasi pembangunan dan kesejahteraan masyarakat. Penelitian ini menerapkan dan mengevaluasi kerangka komputasi berbasis MATLAB untuk estimasi parameter regresi linier berganda dalam menganalisis Produk Domestik Regional Bruto (PDRB) Kota Jayapura. Penelitian menggunakan 15 observasi tahunan periode 2010–2024, dengan investasi infrastruktur, tingkat pengangguran, dan jumlah wisatawan sebagai variabel independen. Prediktor dinormalisasi menggunakan metode Min–Max, kemudian parameter diestimasi dengan algoritma Nelder–Mead melalui fungsi fminsearch dan dibandingkan secara langsung dengan solusi analitik Ordinary Least Squares (OLS). Stabilitas optimisasi dievaluasi menggunakan enam titik awal, sedangkan kemampuan prediksi model dinilai melalui leave-one-out cross-validation. Hasil menunjukkan bahwa OLS dan Nelder–Mead menghasilkan koefisien, nilai prediksi, dan ukuran kinerja yang secara praktis identik, dengan selisih maksimum prediksi sebesar 4.11×10^{-5}. Seluruh titik awal Nelder–Mead berkonvergensi menuju nilai minimum yang sama, dengan rentang Residual Sum of Squares hanya 8.94×10^{-8}. Model menghasilkan R^2=0.7705 dan adjusted R^2=0.7079. Investasi infrastruktur berpengaruh positif dan signifikan terhadap PDRB, sedangkan tingkat pengangguran dan jumlah wisatawan memiliki koefisien negatif tetapi tidak signifikan. Nilai MAPE in-sample sebesar 21.95%, sedangkan MAPE hasil cross-validation sebesar 29.42%, yang menunjukkan kemampuan prediksi moderat. Nilai Durbin–Watson sebesar 0.5644 mengindikasikan autokorelasi positif sehingga hasil inferensi perlu ditafsirkan secara hati-hati. Temuan ini menunjukkan bahwa Nelder–Mead tidak lebih unggul daripada OLS untuk regresi linier tanpa kendala, tetapi mampu mereproduksi solusi OLS secara akurat dan stabil sebagai validasi numerik dalam kerangka komputasi MATLAB.

Kata kunci: Jayapura; MATLAB; Regresi Linier Berganda; Optimisasi Nelder–Mead; Ordinary Least Squares; PDRB.

ABSTRACT

Regional economic performance is an important indicator for evaluating development and public welfare. This study implemented and evaluated a MATLAB-based computational framework for parameter estimation in a multiple linear regression model of Gross Regional Domestic Product (GRDP) in Jayapura. The analysis used 15 annual observations covering the period from 2010 to 2024, with infrastructure investment, unemployment rate, and tourist arrivals as explanatory variables. The predictors were normalized using Min–Max scaling, after which the parameters were estimated using the Nelder–Mead *algorithm through MATLAB's fminsearch function and directly compared with the analytical Ordinary Least Squares (OLS) solution. Optimization stability was evaluated using six independent starting values, while model predictive performance was assessed using leave-one-out cross-validation. The results showed that OLS and Nelder–Mead produced practically identical coefficients, fitted values, and performance measures, with a maximum prediction difference of only 4.11×10^{-5}. All Nelder–Mead starting values converged to essentially the same minimum, with a Residual Sum of Squares range of only 8.94×10^{-8}. The model produced an (R^2) of 0.7705 and an adjusted (R^2) of 0.7079. Infrastructure investment had a positive and statistically significant association with GRDP, whereas unemployment and tourist arrivals had negative but statistically insignificant coefficients. The in-sample MAPE was 21.95%, while the cross-validation MAPE was 29.42%, indicating moderate predictive performance. A Durbin–Watson statistic of 0.5644 indicated positive residual autocorrelation; therefore, the inferential results should be interpreted cautiously. These findings demonstrate that Nelder–Mead does not outperform OLS for unconstrained linear regression, but it can accurately and consistently reproduce the OLS solution as a numerical validation within a MATLAB-based computational framework.*

Keywords: *Jayapura; MATLAB; Multiple Linear Regression; Nelder–Mead Optimization; Ordinary Least Squares; GRDP.*

INTRODUCTION

Regional economic growth serves as a key indicator to evaluate development and public welfare levels within a region. It is commonly reflected by the increase in Gross Regional Domestic Product (GRDP), which captures the dynamics of economic activities, production capacity, and regional resource management effectiveness. Understanding the factors influencing economic growth is essential for formulating sustainable and evidence-based development policies [1]. Among these factors, infrastructure investment, unemployment rates, and tourism sector development are widely recognized as critical drivers of regional economic dynamics. [2], [3].

Empirical studies in Indonesia have consistently shown that these socioeconomic and development-related variables significantly contribute to regional economic performance. For example, a study conducted in East Java using a Seemingly Unrelated Regression (SUR) approach found that investment and several regional development indicators significantly affected economic growth across districts and municipalities [4]. Likewise, previous studies employing panel regression models have demonstrated the effectiveness of statistical approaches in identifying key determinants of regional economic growth and providing evidence to support regional development planning [5]. Similar evidence has also been reported in regional socioeconomic studies using spatial statistical approaches to identify factors affecting regional demographic and development indicators in Indonesia [6].

Jayapura, as a prominent economic hub in eastern Indonesia, exhibits unique characteristics compared to other regions in the country. Local economic growth is shaped by factors such as infrastructure availability, transportation access, service and tourism sector expansion, and labor market conditions. Despite ongoing efforts by the local government to improve infrastructure and boost tourism, challenges such as persistent unemployment and uneven infrastructure development continue to hinder optimal economic growth. [7].

Previous studies in Papua and West Papua have shown that regional development indicators exhibit unique spatial and socioeconomic characteristics compared to other Indonesian regions, emphasizing the need for localized analytical approaches [8]. Furthermore, Recent studies focusing on Papua have utilized time-series forecasting approaches to analyze regional economic growth, highlighting the unique economic characteristics of eastern Indonesia and the need for region-specific analytical models [9].

Although prior studies have explored the relationship between economic factors and GRDP growth, most studies have relied on conventional Ordinary Least Squares (OLS) regression and related econometric approaches to estimate model parameters. OLS provides an analytical solution and remains one of the most widely used techniques for modeling economic relationships. However, the implementation of alternative numerical optimization methods may provide additional flexibility for parameter estimation and serve as a computational validation framework, particularly when extending regression models to more complex optimization settings [10], [11].

Recent developments in computational statistics have encouraged the integration of numerical optimization techniques into regression analysis. One such technique is the Nelder–Mead algorithm, a derivative-free optimization method that has been widely applied in parameter estimation problems across various scientific disciplines [12], [13]. Rather than replacing conventional regression estimation methods, the Nelder–Mead algorithm can be utilized as an alternative numerical approach for minimizing objective functions such as the Residual Sum of Squares (RSS) and validating regression parameter estimates obtained through analytical procedures [14], [15].

In the context of regional economic analysis, studies applying numerical optimization techniques to estimate multivariate economic models remain relatively limited, particularly for regions in eastern Indonesia. Furthermore, empirical studies investigating the combined influence of infrastructure investment, unemployment, and tourism on economic growth in Jayapura are still scarce [16]. Therefore, there is an opportunity to develop and evaluate a MATLAB-based computational framework that integrates multivariate linear regression with numerical optimization techniques for regional economic analysis.

This study aims to implement a MATLAB-based computational framework using the Nelder–Mead optimization algorithm for parameter estimation in a multivariate regression model of regional economic growth in Jayapura. The study examines the effects of infrastructure investment, unemployment rate, and tourism activity on Gross Regional Domestic Product (GRDP). The primary contribution of this research is not the development of a new optimization algorithm, but the demonstration of how numerical optimization techniques can be integrated into a regional economic modeling framework and used as an alternative computational approach for parameter estimation and numerical validation.

By providing an applied case study using economic data from Jayapura, this research contributes to the growing use of computational tools in regional economic analysis and offers a practical framework that may be extended to more complex modeling applications in future studies.

METHOD

Research Type

This study employs a quantitative approach using a multivariate linear regression model to analyze the influence of economic factors on regional economic growth in Jayapura. The quantitative approach was chosen for its ability to provide empirical measurements of relationships among variables through mathematical parameter estimation and statistical analysis. [17]. To enhance the robustness of the analysis, this study integrates a numerical optimization approach using MATLAB for parameter estimation, addressing potential limitations of conventional regression methods.

Data Sources and Data Collection

The dataset used in this study consists of secondary data obtained from official publications by Statistics Indonesia (BPS), local government reports, and other relevant documents on Jayapura's regional economic development. The data spans from 2010 to 2024, allowing for the observation of long-term economic trends. [7]. The dataset consists of annual observations covering the period 2010–2024, resulting in a total of 15 observations. Although the sample size is relatively small, the primary objective of this study is to demonstrate the implementation of a computational parameter estimation framework using available regional economic data rather than to develop a highly generalized predictive model. Nevertheless, the limited number of observations may reduce statistical power and increase the risk of overfitting. Therefore, the results should be interpreted with caution and viewed within the context of exploratory regional economic modeling. The variables included in the analysis are:

- Infrastructure Investment (I): Total investment in infrastructure development, including roads, public facilities, and basic service networks.
- Unemployment Rate (U): The percentage of the working-age population that is unemployed.
- Number of Tourists (W): Annual domestic and international tourist arrivals in Jayapura. Tourism has been identified as a strategic sector contributing to regional economic activity. Previous studies have demonstrated the usefulness of statistical and forecasting models in analyzing tourist arrivals and their economic implications [18].
- Gross Regional Domestic Product (GRDP) (Y): The total economic output of Jayapura, The selection of these variables is based on prior studies that highlight their significant influence on regional economic growth [19]–[22]. However, the study acknowledges the limited sample size (15 annual observations) and its potential impact on statistical power and generalizability. To mitigate this, the analysis emphasizes robust parameter estimation techniques and discusses the implications of the small dataset in the results section. [2], [3].

Multivariate Linear Regression Model

The relationship between the independent variables and the dependent variable is modeled using the following multivariate linear regression equation:

$$Y_t = \beta_0 + \beta_1 I_t + \beta_2 U_t + \beta_3 W_t + \varepsilon_t$$

where:

- Y_t = GRDP value in year t
- I_t = infrastructure investment
- U_t = unemployment rate
- W_t = number of tourists
- β_0 = constant
- $\beta_1, \beta_2, \beta_3$ = regression parameters
- ε_t = error term

The multivariate regression model was selected for its ability to simultaneously evaluate multiple predictors and their combined effects on economic growth [23], [24]. However, the study also considers potential violations of linear regression assumptions, such as multicollinearity, heteroscedasticity, and nonlinearity, which are addressed through diagnostic tests and robust estimation techniques. [25].

Development of the Parameter Estimation Algorithm

Parameter estimation was performed using a numerical optimization approach implemented in MATLAB. In conventional multivariate linear regression, parameter estimates can be obtained analytically through the Ordinary Least Squares (OLS) method. In this study, OLS estimates were first calculated and subsequently used as initial values for the Nelder–Mead optimization algorithm implemented through MATLAB's fminsearch function.

The objective of applying the Nelder–Mead algorithm is not to replace OLS estimation, but rather to provide an alternative computational procedure and numerical validation framework for minimizing the Residual Sum of Squares (RSS). This approach also serves as a foundation for future extensions toward more complex optimization problems where analytical solutions may not be readily available [12]. The parameter estimation process involves the following steps:

1. Initial Estimation: OLS estimates are computed and used as starting values.
2. Objective Function Definition: The Residual Sum of Squares (RSS) is defined as:

$$RSS = \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2$$

3. Numerical Optimization: The Nelder–Mead algorithm iteratively updates parameter values through simplex operations until convergence is achieved.
4. Convergence Assessment: Optimization continues until changes in parameter values and RSS become negligible.
5. Final Estimation: Optimized parameters are compared with OLS estimates to assess consistency and numerical stability.

This procedure enables the evaluation of the computational performance of numerical optimization while maintaining consistency with the underlying linear regression framework [13].

MATLAB Implementation

The computational workflow was implemented using MATLAB R2013a and includes the following stages:

1. Data Preprocessing: Input data is normalized to ensure comparability across variables and to improve optimization stability.
2. Initial Estimation: OLS is used to compute initial parameter values.
3. Optimization: The RSS objective function is minimized using MATLAB's fminsearch function.
4. Model Evaluation: The model's performance is assessed using metrics such as R2, Adjusted R2, F-statistic, and residual analysis.
5. Visualization: Relationships among variables and model predictions are visualized to aid interpretation [26].

Prior to parameter estimation, all predictor variables were normalized using Min–Max normalization to improve numerical stability during optimization. The transformation was performed using:

$$X_{norm} = \frac{(X - X_{min})}{(X_{max} - X_{min})}$$

where X represents the original variable value, Xmin denotes the minimum observed value, and Xmax denotes the maximum observed value. This transformation scales all variables into the interval [0,1], thereby reducing the influence of differences in measurement units and improving optimization convergence. After normalization, the predictor matrix (X) and response variable (Y) were constructed and used in both OLS estimation and Nelder–Mead optimization procedures.

Algorithm Workflow

To enhance clarity, the parameter estimation process is summarized in the following flowchart:

1. Input Data: Load economic data (2010–2024).
2. Preprocessing: Normalize variables and construct predictor (X) and response (Y) matrices.
3. OLS Estimation: Compute initial parameter estimates (β).
4. Objective Function: Define the RSS function.
5. Optimization: Apply Nelder–Mead algorithm using fminsearch.
6. Convergence Check: Repeat iterations until optimal parameters are obtained.
7. Model Evaluation: Compute R2, residuals, and other performance metrics.
8. Output Results: Generate final parameter estimates and visualizations.

Data Analysis Techniques

The data analysis process includes:

1. Descriptive Statistics: Summarize the characteristics of the dataset.
2. Regression Analysis: Estimate parameters using OLS and refine them using Nelder–Mead optimization.
3. Model Diagnostics: Evaluate assumptions (e.g., multicollinearity, heteroscedasticity) and assess model fit.
4. Economic Interpretation: Analyze the influence of each independent variable on GRDP growth.

Comparison of OLS and Nelder–Mead Estimation

To evaluate the computational implementation, parameter estimates obtained using the Nelder–Mead algorithm were compared with those generated by conventional OLS regression. The comparison focused on estimated coefficients, coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and convergence behavior. This comparison was conducted to assess whether the numerical optimization framework produces results consistent with the analytical solution of linear regression.

RESULT AND DISCUSSION

Descriptive Statistics

The analysis was conducted using 15 annual observations covering the period from 2010 to 2024. The descriptive statistics of the variables included in the model are presented in Table 1.

Table 1. Descriptive Statistics of the Research Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
GRDP	14,790.7073	6,632.7395	6,321.4500	25,656.8700
Unemployment Rate	10.8953	1.3357	8.2900	12.7000
Infrastructure Investment	189.6413	35.3687	118.5700	236.5100
Number of Tourists	47.1433	12.0865	28.4100	64.7500

GRDP exhibited considerable variation during the observation period, increasing from a minimum of 6,321.45 to a maximum of 25,656.87. Infrastructure investment and tourist arrivals also showed substantial variation, whereas the unemployment rate had a relatively narrower range.

The limited sample size should be considered when interpreting the results. The model contains three explanatory variables and an intercept estimated from only 15 annual observations. Consequently, hypothesis testing, parameter stability, and predictive generalization should be interpreted cautiously.

Comparison of OLS and Nelder–Mead Parameter Estimates

The parameter estimates obtained using the Nelder–Mead optimization algorithm were compared with those generated by conventional Ordinary Least Squares. OLS was used as an analytical benchmark because the multiple linear regression model has a closed-form least-squares solution. Nelder–Mead was implemented as an alternative numerical procedure for minimizing the Residual Sum of Squares rather than as a method assumed to be superior to OLS. The comparison of the estimated parameters is presented in Table 2.

Table 2. Comparison of OLS and Nelder–Mead Parameter Estimates

Parameter	OLS	Nelder–Mead	Absolute Difference
Intercept	-4,222.601501	-4,222.601413	($8.822\ 10^{-5}$)
Unemployment Rate	-840.508567	-840.508579	($1.248\ 10^{-5}$)
Infrastructure Investment	184.262441	184.262441	($2.820\ 10^{-7}$)
Number of Tourists	-143.664961	-143.664962	($2.396\ 10^{-7}$)

The coefficient estimates produced by the two methods were practically identical. The maximum absolute difference between the estimated coefficients was only 8.822×10^{-5}, while the maximum difference between their predicted values was 4.11×10^{-5}.

These results demonstrate that the Nelder–Mead implementation successfully reproduced the analytical OLS solution with a very high level of numerical precision. However, the findings do not demonstrate that Nelder–Mead improves estimation accuracy or model performance relative to OLS. Instead, the optimization procedure provides a numerical validation of the analytical regression estimates.

The computational contribution of the study therefore lies in demonstrating the implementation and validation of a derivative-free numerical optimization procedure within a regional economic regression framework. This framework may be extended in future research to nonlinear, constrained, or nonstandard objective functions for which analytical solutions are not readily available.

Convergence and Robustness of the Nelder–Mead Algorithm

The convergence and numerical robustness of the Nelder–Mead algorithm were evaluated using six independent starting values. All optimization runs produced an exit flag equal to one, indicating that the convergence criteria were successfully satisfied. The minimum Residual Sum of Squares was: RSS = 141,364,695.3578. The difference between the largest and smallest RSS values across the six starting points was only 8.94 10^{-8}. The number of iterations ranged from 314 to 962, while the number of objective-function evaluations ranged from 619 to 1,692. Although different starting values required different levels of computational effort, all runs converged to essentially the same parameter estimates and objective-function value.

Figure 1 illustrates the differences between the RSS obtained from each starting point and the minimum RSS. The apparent differences in the graph occur only at the order of 10^{-8}. Compared with the total RSS of approximately 1.41×10^{8}, these differences are numerically negligible.

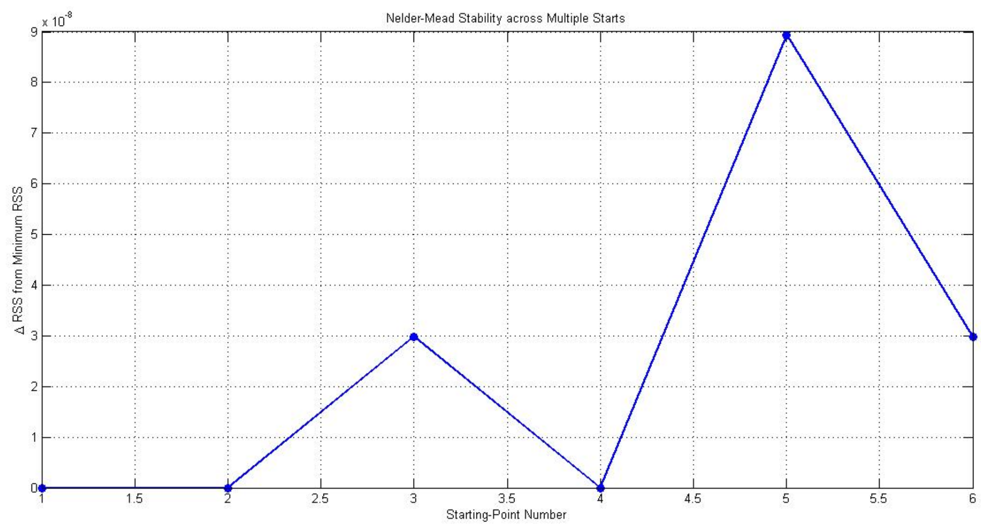

Figure 1. Stability of the Nelder–Mead optimization results across six independent starting values. The vertical axis represents the difference between each RSS value and the minimum RSS.

The multiple-start analysis therefore confirms that the optimization procedure was stable with respect to the selected initial values. This result is also consistent with the convex structure of the least-squares objective function in a full-rank linear regression model, which has a unique global minimum. Accordingly, the multiple-start results demonstrate numerical consistency rather than superiority over OLS.

Regression Model Estimation

Because OLS and Nelder–Mead produced practically identical estimates, the estimated regression model can be expressed as:

-4222.6015 - 840.5086 (Unemployment) + 184.2624 (Investment) - 143.6650 (Tourists).

The inferential statistics were calculated using the conventional OLS covariance matrix. The standard errors, t-statistics, p-values, and confidence intervals are presented in Table 3.

Table 3. Multiple Linear Regression Estimation Results

Variable	Coefficient	Standard Error	t-statistic	p-value	95% Confidence Interval
Intercept	-4,222.6015	10,540.5634	-0.4006	0.6964	[-27,422.2252; 18,977.0222]
Unemployment Rate	-840.5086	739.6367	-1.1364	0.2799	[-2,468.4379; 787.4208]
Infrastructure Investment	184.2624	35.6114	5.1743	0.0003	[105.8823; 262.6426]
Number of Tourists	-143.6650	105.4329	-1.3626	0.2002	[-375.7212; 88.3913]

Infrastructure investment had a positive and statistically significant coefficient. Holding the unemployment rate and tourist arrivals constant, a one-unit increase in infrastructure investment, according to the measurement unit used in the dataset, was associated with an estimated increase of 184.2624 units in GRDP.

The confidence interval for the investment coefficient did not include zero, further supporting its statistically significant positive association with regional economic activity. Among the predictors included in the model, infrastructure investment was the only variable that had a statistically significant partial relationship with GRDP at the 5% significance level.

The unemployment coefficient was negative, indicating that an increase in the unemployment rate tended to be associated with a decrease in GRDP. However, its p-value was greater than 0.05 and its confidence interval included zero. The available data therefore do not provide sufficient evidence that unemployment independently affected GRDP during the observation period.

The coefficient of tourist arrivals was also negative and statistically insignificant after infrastructure investment and unemployment were controlled. This result should not be interpreted as evidence that tourism reduces economic activity. Instead, it indicates that the independent statistical contribution of tourist arrivals could not be established within the current model, sample size, and observation period.

The negative partial coefficient of tourism differs from the positive pattern visible in the bivariate scatter plot. This difference can occur because a scatter plot represents the uncontrolled relationship between two variables, whereas the multiple regression coefficient represents the relationship between tourism and GRDP after the other predictors are held constant.

The discrepancy may also reflect common time trends, omitted economic variables, measurement limitations, or structural changes that are not represented in the current model.

Graphical Relationships Between the Predictors and GRDP

Figure 2 presents the scatter plots between GRDP and each explanatory variable, together with the comparison between actual and OLS-predicted GRDP.

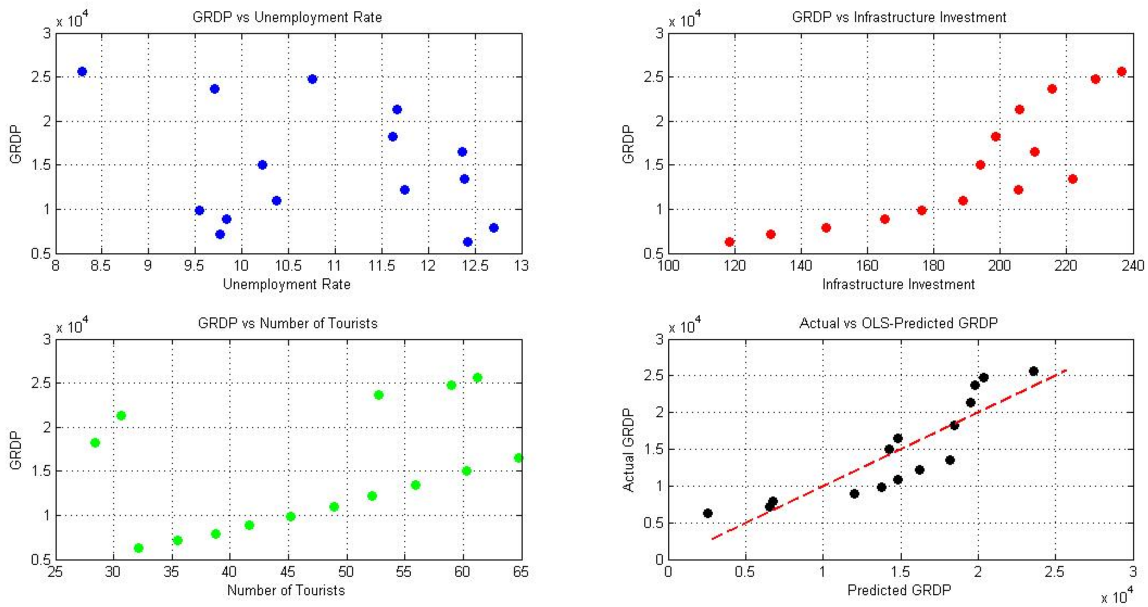

Figure 2. Relationships between unemployment, infrastructure investment, tourist arrivals, and GRDP, together with actual-versus-predicted GRDP values.

The scatter plot between infrastructure investment and GRDP shows the clearest positive relationship among the explanatory variables. Higher infrastructure investment generally corresponds to higher GRDP. This graphical pattern is consistent with the positive and statistically significant investment coefficient. The relationship between unemployment and GRDP is more dispersed. No clear linear pattern is visible across the observations, which is consistent with the statistically insignificant unemployment coefficient.

The scatter plot between tourist arrivals and GRDP displays a generally increasing bivariate pattern. Nevertheless, the tourism coefficient becomes negative and statistically insignificant in the multivariate model. This finding demonstrates that a positive bivariate association does not necessarily imply a positive partial regression effect after other variables are controlled.

The actual-versus-predicted panel shows that most observations are distributed around the 45^{o} reference line, although several observations deviate substantially from it. These deviations indicate that the model captures the general variation in GRDP but does not reproduce all annual observations accurately.

Model Fit and In-Sample Predictive Performance

The model-performance measures obtained from OLS and Nelder–Mead are presented in Table 4.

Table 4. Comparison of OLS and Nelder–Mead Model Performance

Metric	OLS	Nelder–Mead
(R^2)	0.770477	0.770477
Adjusted (R^2)	0.707879	0.707879
SSE	141,364,695.3578	141,364,695.3578
RMSE	3,069.9044	3,069.9044
MAE	2,679.6986	2,679.6986
MAPE	21.9511%	21.9511%

The coefficient of determination indicates that unemployment, infrastructure investment, and tourist arrivals jointly explained approximately 77.05% of the variation in GRDP. After accounting for the number of predictors and the limited sample size, the adjusted coefficient of determination was 0.7079. The overall model was statistically significant, with an F-statistic of 12.3085 and a p-value of 0.000773.

This result indicates that the predictors were jointly associated with GRDP. However, the Mean Absolute Percentage Error of 21.95% shows that the in-sample prediction error was moderate. Therefore, the model should primarily be interpreted as an explanatory regional economic model rather than as a highly accurate forecasting model.

The identical (R^2), adjusted (R^2), SSE, RMSE, MAE, and MAPE values obtained by OLS and Nelder–Mead provide further evidence that both methods converged to the same least-squares solution.

Comparison of Actual, OLS, and Nelder–Mead Estimates

Figure 3 presents the residual diagnostic plots and the annual comparison among actual GRDP, OLS estimates, and Nelder–Mead estimates.

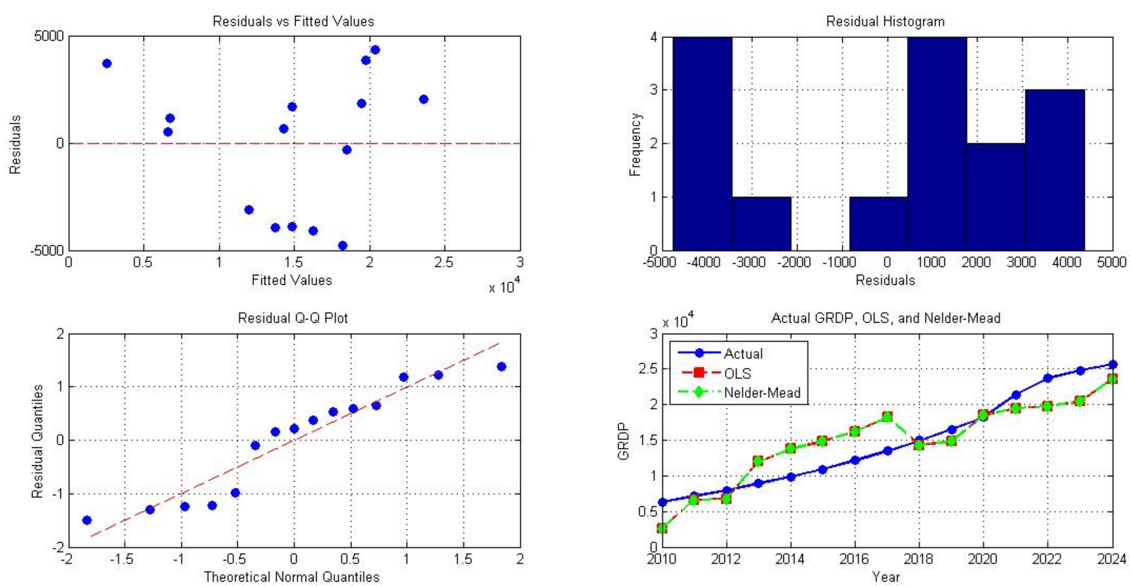

Figure 3. Residual diagnostics and comparison of actual GRDP with OLS and Nelder–Mead estimates for 2010–2024.

The OLS and Nelder–Mead prediction lines overlap almost completely. This visual result confirms that both estimation methods produced practically identical fitted values.

However, the estimated values do not fully follow all annual changes in actual GRDP. During several middle-period observations, the model overestimated GRDP, whereas it underestimated GRDP during several later observations. This pattern suggests that the linear specification may not fully capture the time-dependent development of regional economic activity.

The residual-versus-fitted plot shows residuals on both sides of zero, but their distribution is not entirely random. Several consecutive observations exhibit residuals with similar signs, suggesting the presence of temporal dependence.

The residual histogram is broadly distributed around zero, while the Q-Q plot shows that most residual points approximately follow the theoretical normal reference line. Some deviations remain visible at the distribution tails, which is reasonable given the limited sample size. Therefore, the residual distribution can be described as approximately normal rather than perfectly normal.

Leave-One-Out Cross-Validation

Given the limited sample size, leave-one-out cross-validation was conducted as a sensitivity analysis to evaluate the stability of the model and its ability to predict observations excluded from the estimation process.

The cross-validation results showed that the model achieved a Root Mean Square Error (RMSE) of 4,000.75, a Mean Absolute Error (MAE) of 3,551.50, and a Mean Absolute Percentage Error (MAPE) of 29.42%. In addition, the Predicted Residual Sum of Squares (PRESS) value was 240,089,651.48, while the cross-validated coefficient of determination (Q^2CV) reached 0.6102, indicating a moderate predictive capability of the model.

The LOOCV MAPE was 7.47 percentage points higher than the in-sample MAPE. This increase indicates that the model performs less accurately when applied to observations that were excluded from parameter estimation.

Nevertheless, the positive cross-validated (Q^2) value indicates that the model retains explanatory and predictive information relative to a model that predicts all observations using only the mean GRDP.

Although the Q^2CV value of 0.6102 indicates acceptable predictive relevance, the LOOCV MAPE of 29.42% suggests that the model still produces relatively large prediction errors when predicting unseen observations. This finding indicates that the model is more suitable for explanatory analysis than for high-precision forecasting purposes. The discrepancy between in-sample and cross-validation performance also suggests that some model uncertainty remains when observations are excluded from the estimation process.

The use of cross-validation metrics to assess predictive capability is consistent with previous regression studies that employed Cross Validation and Generalized Cross Validation to evaluate model robustness and forecasting performance [27]. Figure 4 compares the actual GRDP values with the LOOCV predictions.

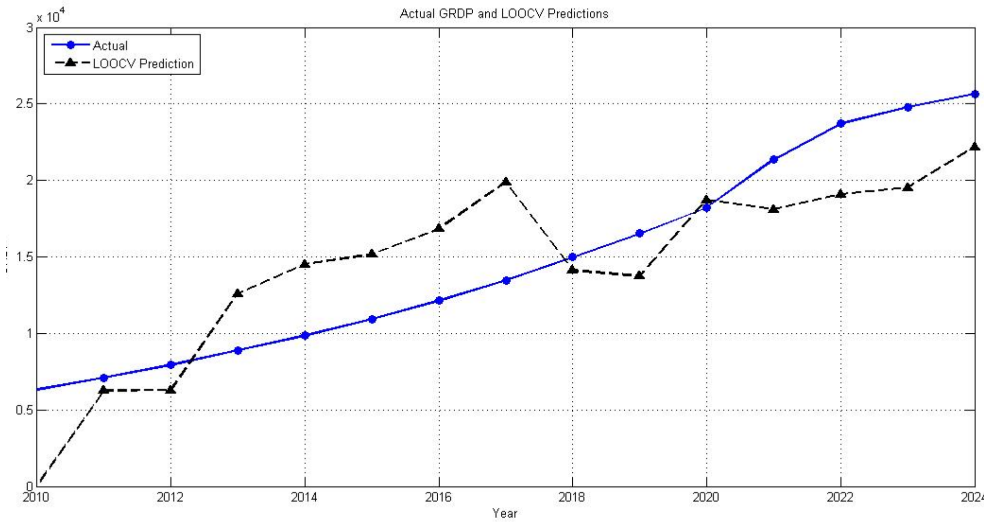

Figure 4. Comparison between actual GRDP and leave-one-out cross-validation predictions.

The largest deviation occurs for the earliest observation. When the first year is omitted, the fitted model must predict a value outside the lower boundary of the remaining GRDP range. This creates an extrapolation problem and produces a relatively large prediction error.

The sensitivity of the first observation is also influenced by the small sample size of only fifteen annual observations. In small datasets, the exclusion of a single observation may substantially affect parameter estimation and prediction performance, leading to larger cross-validation errors than would typically be observed in larger samples.

The LOOCV predictions also deviate from the actual observations during several middle and later years. These deviations are consistent with the higher cross-validation RMSE, MAE, and MAPE values.

Because the data consist of annual time-series observations, LOOCV should be interpreted primarily as a model-stability and sensitivity assessment. It is not a strict simulation of real-time forecasting because later observations may be included when an earlier year is omitted.

Future studies should consider rolling-origin or forward-chaining validation to provide a more appropriate evaluation of temporal forecasting performance. The importance of evaluating predictive performance using validation procedures is consistent with previous studies that emphasized model validation as a critical component of forecasting and prediction accuracy assessment [18].

Regression Diagnostic Tests

The results of the regression diagnostic tests are summarized in Table 5.

Table 5. Regression Diagnostic Test Results

Diagnostic Test	Statistic	p-value	Interpretation
Jarque–Bera	1.2888	0.5250	No evidence against residual normality
Breusch–Pagan	1.4881	0.6850	No evidence of heteroscedasticity
Durbin–Watson	0.5644	—	Indication of positive autocorrelation

The Jarque–Bera test produced a p-value greater than 0.05. Therefore, there was no sufficient statistical evidence to reject the assumption of approximate residual normality. This result is consistent with the general pattern observed in the Q-Q plot.

The Breusch–Pagan test was not statistically significant. Accordingly, the data did not provide evidence of heteroscedasticity, and the residual variance can be considered approximately constant within the limitations of the small sample.

The Variance Inflation Factor values were 1.0633 for unemployment, 1.7282 for infrastructure investment, and 1.7690 for tourist arrivals. All values were below 2, indicating that serious multicollinearity was not present among the explanatory variables.

In contrast, the Durbin–Watson statistic was 0.5644. This relatively low value indicates substantial positive autocorrelation in the residuals. The presence of autocorrelation is important because the data are annual time-series observations and the conventional linear regression model assumes that residuals are independent.

Positive residual autocorrelation does not change the point estimates obtained through OLS, but it may affect the estimated standard errors, t-statistics, confidence intervals, and p-values. Consequently, the statistical significance results should be interpreted cautiously.

Future research should consider autocorrelation-adjusted methods, including Newey–West standard errors, Cochrane–Orcutt estimation, Prais–Winsten regression, or regression models with autoregressive error structures.

Regional Economic Interpretation

The results identify infrastructure investment as the strongest explanatory variable in the model. In the context of Jayapura, infrastructure development may stimulate economic activity through improvements in transportation access, public facilities, market connectivity, production capacity, and the distribution of goods and services. The positive investment coefficient supports the view that infrastructure provision is important for overcoming regional connectivity limitations and supporting economic development in eastern Indonesia [2], [3]. Similar findings have been reported in regional economic modeling studies, where investment-related variables were identified as important determinants of economic growth across Indonesian provinces [4].

The statistically insignificant unemployment coefficient suggests that changes in the unemployment rate alone may not adequately represent the contribution of labor-market conditions to GRDP. The relationship between employment and regional output may also be influenced by informal employment, labor productivity, workforce skills, public-sector employment, and sectoral differences. Similarly, tourist arrivals may not fully reflect the economic value generated by the tourism sector. The number of visitors does not directly measure tourist expenditure, length of stay, hotel occupancy, transportation use, or the amount of income retained by local businesses.

The insignificant tourism coefficient therefore does not imply that tourism is economically unimportant. Instead, it indicates that the independent effect of tourist arrivals was not statistically established after investment and unemployment were controlled. Future studies should include more direct tourism indicators, such as average tourist expenditure, hotel occupancy rates, length of stay, tourism-sector employment, and tourism-related regional revenue.

Computational Interpretation

The direct comparison between OLS and Nelder–Mead addresses the computational objective of the study. The Nelder–Mead estimates reproduced the analytical OLS estimates with extremely small numerical differences. The multiple-start results also demonstrated that all optimization runs converged to the same minimum, even though the number of iterations differed across initial values.

These findings confirm that the MATLAB implementation was numerically accurate and stable for the analyzed dataset. However, the results do not support the claim that Nelder–Mead is more accurate, more robust, or more effective than OLS for an unconstrained linear regression model.

The role of Nelder–Mead in this study is therefore best described as an alternative numerical estimation procedure and a validation of the analytical regression solution. Its broader relevance lies in the possibility of extending the same computational framework to nonlinear regression, constrained parameter estimation, robust loss functions, or other regional economic models for which closed-form analytical solutions are unavailable.

Study Limitations and Implications

Several limitations should be considered when interpreting the findings. First, only 15 annual observations were available. This limits statistical power, increases uncertainty in coefficient estimates, and restricts the generalizability of the results. Second, the Durbin–Watson statistic indicates substantial positive residual autocorrelation. Therefore, the conventional standard errors and hypothesis tests should be interpreted cautiously. Third, the model includes only unemployment, infrastructure investment, and tourist arrivals. Other determinants of regional economic activity, including government expenditure, inflation, population, education, trade activity, household consumption, and sectoral production, were not included. Fourth, the cross-validation results show that predictive errors increase when individual observations are excluded from model estimation. The model should therefore be considered an applied explanatory framework rather than a high-accuracy forecasting system.

Despite these limitations, the study provides evidence that infrastructure investment is positively associated with GRDP and demonstrates that a MATLAB-based Nelder–Mead procedure can numerically reproduce the analytical OLS solution. Future research should use higher-frequency data, include additional macroeconomic indicators, correct for residual autocorrelation, and apply time-series validation procedures.

CONCLUSION

This study implemented a MATLAB-based computational framework for estimating a multivariate linear regression model of regional economic growth in Jayapura using Ordinary Least Squares and the Nelder–Mead optimization algorithm. The analysis was based on 15 annual observations covering the period from 2010 to 2024.

The empirical results indicate that infrastructure investment had a positive and statistically significant association with Gross Regional Domestic Product. In contrast, the unemployment rate and the number of tourists had negative but statistically insignificant coefficients after the other explanatory variables were controlled. These findings suggest that infrastructure investment was the strongest explanatory variable in the estimated model.

The regression model explained approximately 77.05% of the variation in GRDP, with an adjusted coefficient of determination of 70.79%. The model was statistically significant as a whole. However, the in-sample MAPE of 21.95% and the LOOCV MAPE of 29.42% indicate moderate predictive accuracy. Therefore, the model should primarily be interpreted as an explanatory framework rather than as a highly accurate forecasting model.

The cross-validation results further indicate that prediction performance decreases when observations are omitted from model estimation, suggesting that caution is required when applying the model for forecasting purposes beyond the observed sample period.

The comparison between OLS and Nelder–Mead showed that both methods produced practically identical coefficient estimates, fitted values, and model-performance measures. The maximum difference between their predicted values was only $4.11\ 10^{-5}$. In addition, all six independent Nelder–Mead starting values converged to essentially the same minimum RSS. These results confirm the numerical accuracy and stability of the MATLAB implementation.

Nevertheless, the findings do not demonstrate that Nelder–Mead is superior to OLS for an unconstrained linear regression model. In this study, the optimization algorithm served as an alternative numerical estimation procedure and as a validation of the analytical OLS solution. Its broader potential lies in its possible extension to nonlinear, constrained, or nonstandard optimization problems for which closed-form solutions are unavailable.

The diagnostic results showed no evidence of serious multicollinearity, heteroscedasticity, or substantial departure from residual normality. However, the Durbin–Watson statistic indicated positive residual autocorrelation. Consequently, the conventional OLS standard errors, confidence intervals, and significance tests should be interpreted cautiously.

The study is also limited by the small number of annual observations and the exclusion of other potentially relevant economic variables. Future research should use higher-frequency data, incorporate additional determinants of regional economic growth, apply autocorrelation-adjusted inference, and employ rolling-origin or forward-chaining validation. These improvements would provide a more robust assessment of both the explanatory and predictive performance of regional economic models.

REFERENCE

[1] Y. Jin, Y. Ge, H. Fan, Z. Li, Y. Liu, and Y. Jia, "Mapping Gross Domestic Product Distribution at 1 km Resolution across Thailand Using the Random Forest Area-to-Area Regression Kriging Model," *Int. J. Geo-Information*, vol. 12, no. 12, p. 481, 2023, doi: https://doi.org/10.3390/ijgi12120481.

[2] M. M. Alalawneh, J. Mammadov, and A. Alqasem, "Nexus between FDI , Infrastructure Investment , Tourism Revenues , and Economic Growth : Mega Event Evidence," *Emerg. Sci. J.*, vol. 5, no. 6, pp. 953–963, 2021, doi: https://doi.org/10.28991/esj-2021-01323.

[3] R. Karim *et al.*, "ECONOMIC CONTRIBUTION OF TOURISM IN GILGIT-BALTISTAN, PAKISTAN," *J. Malaysian Inst. Planners*, vol. 21, no. 4, pp. 302–309, 2023, doi: https://doi.org/10.21837/pm.v21i28.1334.

[4] A. Fitriana, W. Pramesti, and E. M. P. Hermanto, "Pemodelan Seemingly Unrelated Regression (Sur) Pada Faktor Pertumbuhan Perekonomian Di Provinsi Jawa Timur Tahun 2018," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 13, no. 2, pp. 12–19, 2020, doi: 10.36456/jstat.vol13.no2.a2950.

[5] M. A. Abdullah and Asrirawan, "Model Pertumbuhan Ekonomi Kabupaten Kota Sulawesi Barat Menggunakan Pendekatan Regresi Spasial Data Panel," *Saintifik J. Mat. Sains, dan Pembelajarannya*, vol. 7, no. 1, pp. 34–44, 2021, doi: 10.31605/saintifik.v7i1.298.

[6] S. Wahyuni, N. Amalita, and F. Fitri, "Implementation of Spatial Autoregressive Analysis to Determine Factors Affecting the Population Dependency Ratio in West Sumatera," *J Stat.*, vol. 18, no. 2, pp. 957–967, 2025.

[7] J. C. Karay, F. X. Sugiyanto, and W. Widodo, "International Journal of Sustainable Development and Planning The Effects of Electricity Consumption on Electrification Access on Economic Growth in Papua Province , Indonesia," vol. 17, no. 6, pp. 1747–1752, 2022.

[8] D. S. Hastuti and S. Yulianto, "Factors Affecting the Resilience Index Food in Papua Province and West Papua Province Using a Spatial Model Approach," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 17, no. 1, pp. 635–644, 2024, doi: 10.36456/jstat.vol17.no1.a9087.

[9] B. F. Kuddi and F. Seru, "Peramalan pertumbuhan Ekonomi di Provinsi Papua Menggunakan Model Hybrid Arima-Garch," *Saintifik J. Mat. Sains, dan Pembelajarannya*, vol. 11, no. 1, pp. 104–109, 2025, doi: 10.31605/saintifik.v11i1.578.

[10] W. Albra, T. Zulham, I. Majid, B. Subartini, and F. Thalia, "The effect of gross domestic product and population growth on CO2 emissions in Indonesia: An application of the ant colony optimisation algorithm and cobb-douglas model," *Int. J. Energy Econ. Policy*, vol. 9, no. 4, pp. 313–319, 2019.

[11] S. Alfiani and P. R. Arum, "Pemodelan Pertumbuhan Ekonomi di Jawa Barat Menggunakan Metode Geographically Weighted Panel Regression Universitas Muhammadiyah Semarang," *J Stat.*, vol. 15, no. 2, pp. 219–227, 2022.

[12] K. Kumar and E. Kostina, "Optimal Parameter Estimation Techniques for Complex Nonlinear Systems," *Differ. Equations Dyn. Syst.*, vol. 34, no. 2, pp. 597–616, 2026, doi: 10.1007/s12591-024-00688-9.

[13] M. Németh, J. Hegedüs, G. Hantos, A. K. Abdulrazzaq, and A. Poppe, "Parameter Extraction for a SPICE-like Delphi4LED Multi-Domain Chip-Level LED Model with an Improved Nelder – Mead Method," *Appl. Sci.*, vol. 14, no. 16, p. 7186, 2024, doi: https://doi.org/10.3390/app14167186.

[14] A. M. Astuti, Setiawan, I. Zain, and J. D. T. Purnomo, "A modified generalized estimating equation approach for simultaneous spatial durbin panel model: Case study of economic growth in ASEAN countries," *Decis. Sci. Lett.*, vol. 12, no. 2, pp. 369–388, 2023, doi: https://doi.org/10.5267/j.dsl.2023.1.001.

[15] Z. Wang *et al.*, "Exploring the Potential of OpenStreetMap Data in Regional Economic Development Evaluation Modeling," *Remote Sens. Artic.*, vol. 16, no. 2, p. 239, 2024, doi: https://doi.org/10.3390/rs16020239.

[16] T. S. Guntoro *et al.*, "The design of contextual domain tourism sports through traditional sports in jayapura indonesia El diseño del dominio contextual del turismo deportivo por parte

de los deportes tradicionales en Jayapura, Indonesia," *Retos*, vol. 52, pp. 164–170, 2024, doi: https://doi.org/10.47197/retos.v52.101626.

[17] G. Anuraga, A. Indrasetianingsih, and M. Athoillah, "Pelatihan pengujian hipotesis statistika dasar dengan software r," BUDIMAS: Jurnal Pengabdian Masyarakat, vol. 3, no. 2, pp. 327–334, 2021.

[18] Evita Purnaningrum and H. K. Nafah, "Analisis Kalman filter berbasis Google Trends untuk Prediksi Kedatangan Wisatawan Mancanegara di Indonesia Pasca Pandemi," *J Stat.*, vol. 14, no. 2, pp. 110–115, 2022, doi: 10.36456/jstat.vol14.no2.a4956.

[19] G. Theofani and E. Sediyono, "Multiple Linear Regression Analysis on Factors that Influence Employees Work Motivation," *Sink. J. dan Penelit. Tek. Inform.*, vol. 6, no. 3, pp. 791–798, 2022, doi: https://doi.org/10.33395/sinkron.v7i3.11453 e-ISSN.

[20] K. Muzzakar, S. Syahnur, and M. Abrar, "Provincial Real Economic Growth in Indonesia : Investigating Key Factors Through Spatial Analysis," *Ekon. J. Econ.*, vol. 1, no. 2, pp. 40–50, 2023, doi: 10.60084/eje.v1i2.66.

[21] J. Jasman and H. Hwihanus, "Economic Growth Dynamics in West Nusa Tenggara Province: Investment, Gross Regional Domestic Product, and Government Finance," *J. Ilm. Manaj. Kesatuan*, vol. 11, no. 3, pp. 905–918, 2023.

[22] A. N. Putri, D. Hariyanti, and M. Adriana, "The Influence Of Local Own-Source Revenue, Human Development Index, And Labor Force Participation Level On Grdp Per Capita Moderate By The Depth Of Poverty In City Districts Of Papua Province In 2017-2021," *Indones. Interdiscip. J. Sharia Econ.*, vol. 7, no. 2, pp. 2895–2909, 2024.

[23] M. K. Ardakani and S. M. Seyedaliakbar, "Impact of energy consumption and economic growth on CO2 emission using multivariate regression," *Energy Strateg. Rev.*, vol. 26, no. November, p. 100428, 2019, doi: 10.1016/j.esr.2019.100428.

[24] M. Bouznit, M. P. Pablo-Romero, and A. Sánchez-Braza, "Economic growth, human capital, and energy consumption in Algeria: evidence from cointegrating polynomial regression and a simultaneous equations model," *Environ. Sci. Pollut. Res.*, vol. 30, no. 9, pp. 23450–23466, 2023, doi: 10.1007/s11356-022-23657-7.

[25] R. Zhao, X. Gu, B. Xue, J. Zhang, and W. Ren, "Short period PM2.5 prediction based on multivariate linear regression model," *PLoS One*, vol. 13, no. 7, pp. 1–15, 2018, doi: https://doi.org/10.1371/journal.pone.0201011.

[26] T. Storcz, I. Kistelegdi, K. R. Horváth, and Z. Ercsey, "Applicability of Multivariate Linear Regression in Building Energy Demand Estimation," *Math. Model. Eng. Probl.*, vol. 9, no. 6, pp. 1451–1458, 2022, doi: https://doi.org/10.18280/mmep.090602.

[27] G. K. Ma'rifah, M. Idhom, and Trimono, "Implementation of Geographically and Temporally Weighted Regression with Cross Validation and Generalized Cross Validation Methods for Deforestation Modeling in Kalimantan," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 18, no. 1, pp. 840–850, 2025, doi: 10.36456/jstat.vol18.no1.a10331.