

Correlation Analysis of the Effectiveness of Electric Energy Production at PT PLN (Persero) based on Panel Data Regression Method

Agista Surya Bawana¹, Nurcahya Yulian Ashar², Humam Rosyadi³,
Arya Wicaksana⁴

^{1,2,4}Department of Mathematics, Faculty of Sciences and Mathematics, Diponegoro University, Semarang, Indonesia.

³Department of Accounting, Politeknik Negeri Semarang, Semarang, Indonesia.

Corresponding Author: Nurcahya Yulian Ashar

DOI: <https://doi.org/10.52403/ijrr.20251239>

ABSTRACT

This study aims to analyze the correlation between the effectiveness of electricity production at PT PLN (Persero) using panel data regression methods. Given the high demand for electricity in line with economic growth and population growth, this state-owned electricity company needs to ensure efficiency in its production. This study uses secondary data from the 2023 PT PLN Statistical Report to evaluate the influence of variables such as load factors, capacity, energy sold per customer group, energy produced, and the number of generating units on company revenue. The methodology used is multiple linear regression with a panel data approach, as well as a series of statistical tests to ensure data validity, such as normality tests, multicollinearity tests, and Durbin-Watson autocorrelation tests. The analysis results show that only one variable, namely energy sold per customer group, has a significant effect on company revenue. Simultaneously, the five variables tested contributed 74.3% to the company's revenue. Based on these findings, it is recommended that PT PLN (Persero) focus more on increasing energy sales per customer group, particularly in the business and industrial sectors. Increased

promotion and customer acquisition programs can be effective strategies for increasing revenue and production efficiency, as well as having a positive impact on the company's profit margin.

Keywords: Panel Data Regression, Electric Energy Production, Effectiveness Analysis, Energy Efficiency

INTRODUCTION

Electrical energy constitutes a foundational pillar of national infrastructure, supporting economic productivity, industrial operations, public services, and household activities. Indonesia's increasing electrification ratio and the expansion of economic activity have driven a steady rise in electricity demand in recent years [1][2]. As the national electricity provider, PT PLN (Persero) is therefore required to sustain reliable electricity performance while simultaneously ensuring cost efficiency and long-term financial viability.

Understanding the determinants of production effectiveness is essential for optimizing revenue generation and resource utilization. Operational indicators—including load factor, capacity factor, energy generated, and the number of power plants—serve as measures of system

efficiency and reliability within the electricity supply chain [3][4]. At the same time, energy sold per customer group directly reflects consumer behavior and revenue-forming demand patterns, making it a central variable in utility finance [5][6].

Although these variables are widely acknowledged in energy economics, research focusing on their comparative influence on PLN's revenue using structured econometric modeling remains limited. Accordingly, this study aims to examine the significance of key operational variables in determining PLN's revenue by employing multiple linear regression complemented by classical assumption tests [7][8].

This study contributes to the field of energy management and public-utility optimization by providing evidence-based insights that can support PLN in strengthening operational planning, financial forecasting, and strategic decision-making.

LITERATURE REVIEW

A. Energy Demand and Economic Activity

Electricity consumption is strongly associated with national economic performance. As industrial activities expand and household appliance usage increases, electricity demand typically exhibits upward trends. Data from the Ministry of Energy and Mineral Resources indicate a consistent rise in electricity demand, signaling improvements in economic capacity and public welfare [9].

B. Multiple Linear Regression

Multiple Linear Regression (MLR) is a statistical method used to analyze the relationship between one dependent variable and multiple independent variables [10]. Valid estimates require compliance with classical assumptions such as normality, homoscedasticity, and independence of residuals [11].

The general model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

C. Panel Data Regression

Panel data integrate cross-sectional and time-series dimensions, improving estimation accuracy and controlling for unobserved heterogeneity [12][13]. Although the present research uses cross-sectional data, the conceptual framework aligns with the principles of panel-data modeling, particularly regarding the treatment of regional variation within PLN's operational units [14].

D. Classical Assumption Tests

From [15], regression validity is assessed through several diagnostics:

1. Normality test (Kolmogorov–Smirnov).
2. Multicollinearity test (VIF and Tolerance).
3. Autocorrelation test (Durbin–Watson).
4. t-test for partial significance.
5. F-test for simultaneous significance.

MATERIALS & METHODS

A. Research Design

This study adopted a quantitative descriptive research design, aiming to systematically evaluate the relationship between operational indicators and revenue performance at PT PLN (Persero). This quantitative approach is appropriate because the study examines measurable numerical variables and attempts to estimate the magnitude of the influence between predictors and dependent variables using statistical modeling. The descriptive component of this design aims to characterize the patterns, structures, and relationships between variables in PLN's operational and financial data.

This analysis utilizes a multiple linear regression framework, supported by classical assumption tests, to ensure the accuracy, validity, and reliability of the estimated parameters. This methodological orientation enables this study to generate evidence-based insights that can be applied to operational planning and strategic decision-making in the electricity sector.

B. Data Source

This study uses secondary data obtained from the 2023 PLN Statistical Report, officially published by PT PLN (Persero).

This document contains comprehensive annual operational and financial statistics covering various aspects of electricity generation, distribution, and consumption. This dataset consists of observations from 23 regional operating units, representing PLN's administrative divisions across Indonesia. These units serve as a cross-cutting element of this analysis, providing a diverse representation of geographic, demographic, and operational conditions across the national electricity grid.

C. Variables

This study includes one dependent variable and five independent variables, each selected based on its theoretical relevance to revenue generation in electricity utilities.

1. Dependent Variable

a. Revenue (%)

Represents the proportion or level of revenue generated by each PLN regional operating unit. Revenue is influenced by demand-side factors such as electricity consumption and customer mix, as well as supply-side operational performance.

2. Independent Variable

a. Load Factor (%)

Measures the utilization rate of installed electricity capacity over a specified period. A higher load factor indicates more consistent usage relative to peak demand.

b. Capacity Factor (%)

Represents the ratio of actual electricity generated to the maximum generation possible if all power plants were operating at full capacity continuously. This indicator reflects operational efficiency and production reliability.

c. Energy Sold per Customer Group (%)

Includes the proportion of electricity sales distributed across all customer segments (residential, commercial, industrial, public services, and others). This variable serves as a key demand-side indicator closely related to revenue generation.

d. Energy Generated (GWh)

Refers to the total volume of electricity generated by PLN's generating units, expressed in gigawatt-hours. This variable

reflects the supply-side capacity of the electricity system.

e. Number of Power Plants (units)

Indicates the total number of generating facilities operating in each region. This structural variable captures infrastructure availability and potential production capacity.

D. Data Processing Steps

The data analysis procedure consists of several structured steps designed to ensure methodological rigor and the reliability of the statistical results:

1. Data Extraction

Relevant operational and financial indicators were extracted from the 2023 PLN Statistics dataset. Only variables aligned with the study's analytical framework were selected for further processing. Data Standardization and Preprocessing

Data were checked for completeness, consistency, and uniformity. Measurement units were standardized where necessary, and missing or anomalous entries were handled to maintain dataset integrity.

2. Normality of residuals using the Kolmogorov–Smirnov test

Multicollinearity assessment using the Variance Inflation Factor (VIF) and tolerance values.

3. Autocorrelation detection using the Durbin–Watson statistic

These tests ensure that the model produces valid, unbiased, and efficient estimators.

4. Multiple Linear Regression Analysis

Regression modeling was performed using SPSS software, which allows for the estimation of coefficients describing the influence of each independent variable on the dependent variable. This technique quantifies the extent to which operational and demand-side factors shape regional revenue outcomes.

5. Model Evaluation and Interpretation

Model performance and variable significance were evaluated using:

- a. t-test to assess the partial significance of each independent variable.

- b. F-test to determine the overall significance of the model Coefficient of Determination (R^2) for m.

RESULT

A. Normality Test

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Faktor_Beban	.127	23	.200 [*]	.948	23	.270
Faktor_Kapasitas	.126	23	.200 [*]	.979	23	.885
Energi_Terjual_perKelompok_Pelanggan	.299	23	.063	.703	23	.087
Energi_yang_diproduksi	.226	23	.053	.801	23	.074
Jumlah_Unit_Pembangkit	.258	23	.070	.824	23	.081

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

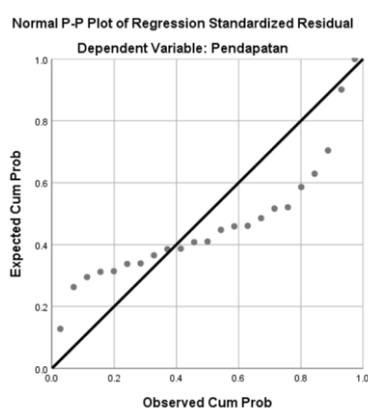


Figure 1. Test of Normality and Normal P Output

The Kolmogorov–Smirnov normality test shows that all variables used in the regression model have a significance value greater than 0.05, confirming that the dataset follows a normal distribution. This result is important because the linear regression framework assumes that residuals come from a normally distributed population. Meeting this assumption ensures that subsequent statistical inferences, such as hypothesis testing and confidence interval estimation, remain valid and reliable.

In addition to the numerical tests, graphical diagnostics including a P plot and a residual histogram indicate that the residual distribution closely conforms to the expected normal curve. The points in the P plot follow the diagonal line, indicating minimal deviation from normality. This visual confirmation further strengthens the validity of the regression model and

supports its suitability for analytical interpretation.

B. Multicollinearity Test

Model	Coefficients ^a				Sig.	Collinearity Statistics	
	Unstandardized Coefficients	Standardized Coefficients	t	Sig.		Tolerance	VIF
1 (Constant)	202.903	266.575	.761	.457			
Faktor_Beban	.014	.040	.052	.356	.726	.717 1.395	
Faktor_Kapasitas	-.018	.063	-.052	-.294	.772	.483 2.069	
Energi_Terjual_perKelompok_Pelanggan	.982	.178	.787	5.509	.000	.741 1.350	
Energi_yang_diproduksi	.000	.001	-.032	-.171	.867	.433 2.307	
Jumlah_Unit_Pembangkit	-.351	.330	-.164	-1.064	.302	.633 1.580	

a. Dependent Variable: Pendapatan

Figure 2. Multicollinearity Test Output

The multicollinearity assessment showed that all independent variables had tolerance values above 0.1 and VIF values below 10, indicating the absence of problematic multicollinearity. This indicates that the predictors are not highly correlated with each other and can independently contribute to explaining variation in the dependent variable. Consequently, the regression coefficients generated by the model remain stable and interpretable.

The absence of multicollinearity is crucial to ensuring the integrity of a regression analysis. Multicollinearity can increase standard errors, weaken the significance of predictors, and distort coefficient estimates. Because the current model does not suffer from these issues, each variable can be evaluated on its own merits, allowing for a more accurate interpretation of the underlying relationships.

C. Autocorrelation Test

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.862 ^a	.743	.667	312.71162	2.835

a. Predictors: (Constant), Jumlah_Unit_Pembangkit, Faktor_Kapasitas, Energi_Terjual_perKelompok_Pelanggan, Faktor_Beban, Energi_yang_diproduksi

b. Dependent Variable: Pendapatan

Figure 3. Autocorrelation Test Output

The Durbin–Watson statistic of 2.835 indicates negative autocorrelation among the residuals. Although this study used cross-sectional data, this phenomenon likely arises from structural similarities across PLN's regional units. These similarities likely stem from uniform regulatory practices, standardized operational procedures, and consistent tariff schemes applied across regions, leading to patterns in residual behavior.

While negative autocorrelation does not necessarily invalidate the regression model, it does imply that the residuals do not vary randomly across units. This may indicate the presence of underlying patterns influenced by organizational or geographic uniformity. Future studies are recommended to use time series or panel data to analyze this correlation structure and obtain more accurate estimates.

D. Partial Significance (t-Test)

Model	Coefficients ^a						Collinearity Statistics	
	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	Tolerance	VIF		
1 (Constant)	202.903	.266	.575	.761	.457			
Faktor_Beban	.014	.040	.052	.356	.726	.717	1.395	
Faktor_Kapasitas	-.018	.063	-.052	-.294	.772	.483	2.069	
Energi_Terjual_perKelompok_Pelanggan	.982	.178	.787	5.509	.000	.741	1.350	
Energi_yang_diproduksi	.000	.001	-.032	-.171	.867	.433	2.307	
Jumlah_Unit_Pembangkit	-.351	.330	-.164	-1.064	.302	.633	1.580	

a. Dependent Variable: Pendapatan

Figure 4. t-Test Output

A partial significance test indicates that only the Energy Sold per Customer Group variable has a statistically significant effect on revenue ($p = 0.000$). This finding reinforces the inherent structure of the electricity industry, where revenue is directly related to the volume of electricity sold. Higher energy consumption levels reflect increased customer activity and greater reliance on PLN's electricity supply, thus driving financial performance. In contrast, the other operational variables—load factor, capacity factor, energy

generated, and number of power plants—did not show significant individual effects. This suggests that these indicators primarily measure operational efficiency rather than revenue generation. Their effect on revenue is indirect and likely dependent on broader demand-side factors, which explains the lack of partial significance in the model.

E. Simultaneous Significance (F-Test)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4800377.800	5	960075.560	9.818	.000 ^b
	Residual	1662405.504	17	97788.559		
	Total	6462783.304	22			

a. Dependent Variable: Pendapatan

b. Predictors: (Constant), Jumlah_Unit_Pembangkit, Faktor_Kapasitas

Figure 5. F-Test Output

The F-test yielded a significance value of 0.000, indicating that all independent variables jointly influence PLN's revenue. Although some variables were not individually significant, their combined effect was substantial, highlighting the interconnected nature of operational and demand-driven factors in the electricity sector. This simultaneous impact validates the regression model and emphasizes the multidimensional determinants of financial results.

The overall model significance indicates that operational efficiency, production capacity, and electricity sales collectively shape revenue performance. This underscores the need to evaluate these factors in an integrated manner, rather than separately. Therefore, strategic planning at PLN must consider both supply- and demand-driven dynamics to optimize revenue generation.

F. Coefficient of Determination

Model Summary ^b					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.862 ^a	.743	.667	312.71162	2.835

a. Predictors: (Constant), Jumlah_Unit_Pembangkit, Faktor_Kapasitas

Figure 6. Coefficient of Determination Output

The R-squared value of 0.743 indicates that 74.3% of the revenue variation can be explained by the five predictor variables in

the model. This reflects high analytical power, indicating that the selected variables are highly relevant to understanding revenue fluctuations across PLN's regional units. Therefore, this model provides a reliable representation of the factors influencing financial performance.

The remaining 25.7% of unexplained variance likely stems from external influences not included in the model.

DISCUSSION

The identification of Energy Sold per Customer Group as the only individually significant predictor aligns with the fundamental revenue structure of electric utility enterprises, where financial performance is primarily driven by the volume of electricity billed to consumers. This variable reflects actual market demand and customer utilization patterns, which are shaped by industrial activity, demographic conditions, and broader macroeconomic dynamics. Consequently, increases in electricity sales indicate stronger economic engagement and directly enhance revenue streams, reaffirming the centrality of consumption-driven indicators in determining PLN's financial sustainability. Operational indicators such as load factor and capacity factor did not exhibit significant partial effects because these metrics principally represent internal system efficiency rather than direct revenue determinants. Although higher efficiency can reduce operational costs, it does not automatically translate into increased electricity utilization or higher sales volumes. Previous studies similarly suggest that operational efficiency contributes meaningfully to profitability only when supported by strong consumer demand and utilization growth.

Thus, while enhancing operational efficiency remains strategically important, it must be complemented by demand-oriented initiatives to yield significant revenue impacts.

The detection of negative autocorrelation within the cross-sectional dataset suggests a

high level of structural homogeneity among PLN's regional units. As a centrally regulated state-owned enterprise, PLN operates under standardized tariff policies, uniform technical protocols, and consistent performance benchmarks across regions. These institutional uniformities contribute to similar behavioral patterns among units, resulting in non-random residual configurations. Recognizing this phenomenon is essential for improving future empirical models, especially those seeking to incorporate region-specific heterogeneity or spatial dependence.

The simultaneous significance of all variables highlights that operational readiness and consumption dynamics operate collectively within an integrated system influencing revenue outcomes. This indicates the need to harmonize supply-side capabilities—such as generation capacity, infrastructure adequacy, and energy availability—with demand-side factors including customer satisfaction and consumption behaviors. Therefore, neither operational efficiency nor market demand alone can ensure sustainable revenue growth; a comprehensive, system-wide approach is required.

From a managerial perspective, the findings underscore the importance of prioritizing strategies that stimulate electricity utilization, particularly within high-demand sectors such as industry, commerce, and public services. Enhancing service reliability, accelerating electrification programs, improving customer relations, and fostering electricity-intensive economic activities can significantly strengthen revenue performance.

For future scientific research, the use of longitudinal panel data, integration of tariff-related variables, or utilization of spatial econometric techniques would yield deeper insights into temporal and regional variations, thereby improving predictive accuracy and broadening policy relevance.

CONCLUSION

The results of the analysis of PT PLN (Persero) statistical data for 2023 using the Correlation and Multiple Linear Regression Analysis method show that the model used has a strong ability to explain the relationship between the research variables. The R-Square value of 74.3% indicates that the five independent variables simultaneously have a significant influence on PLN's revenue. This indicates that the regression model constructed is suitable for use as a basis for understanding the factors that play a role in determining the effectiveness of electricity production.

In the partial analysis, only the variable Energy Sold per Customer Group was proven to have a significant effect on PLN's revenue, with a significance value of less than 0.05. This finding shows that the main factor determining the company's revenue is the amount of electricity successfully sold to customers. Meanwhile, other variables such as Load Factor, Capacity Factor, Energy Produced, and Number of Generating Units did not show a significant partial effect, although they still contributed when tested simultaneously.

This study confirms that efforts to increase PT PLN (Persero)'s revenue can be focused on strategies that encourage an increase in the amount of energy sold to customers. Optimizing energy distribution, improving service quality, and innovating in marketing and customer management are important steps that can increase the effectiveness of electricity production. Thus, the results of this study are expected to serve as a reference for the company in formulating more efficient operational policies and supporting sustainable profit growth.

Declaration by Authors

Acknowledgement: None

Source of Funding: None

Conflict of Interest: No conflicts of interest declared.

REFERENCES

1. Girik Allo, A., Dwiputri, I. N., & Maspaitella, M. (2022). The impact of electricity investment on inter-regional economic development in Indonesia: An Inter-Regional Input-Output (IRIO) approach. *Journal of Socioeconomics and Development*, 5(1), 1–12. <https://doi.org/10.31328/jsed.v5i1.2775>
2. Karay, J.C., Firmansyah, Sugiyanto, F.X., Widodo, W. (2022). The effects of electricity consumption on electrification access on economic growth in Papua Province, Indonesia. *International Journal of Sustainable Development and Planning*, Vol. 17, No. 6, pp. 1747-1752. <https://doi.org/10.18280/ijstdp.170608>
3. Najjar, Y.S.H., Abu-Shamleh, A. Performance evaluation of a large-scale thermal power plant based on the best industrial practices. *Sci Rep* 10, 20661 (2020). <https://doi.org/10.1038/s41598-020-77802-8>
4. D. A. Ekpechi, "Evaluation of Key Performance Factors and Recommendation of Optimization Strategies of a Power Generation Company", *Engineering Science & Technology*, vol. 6, no. 1, pp. 52–68, Nov. 2024
5. P. Wang, D. Fang, and G. C. Cao, "How social learning affects customer behavior under the implementation of TOU in the electricity retailing market," *Energy Economics*, vol. 106, 2022, doi: 10.1016/j.eneco.2022.105836
6. T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," *Applied Energy*, vol. 87, no. 11, 2010, doi: 10.1016/j.apenergy.2010.05.015.
7. Indriakati A. J., Zulfayani A., & Siska V. (2022). Analisis Pengaruh Biaya Operasional Terhadap Kinerja Keuangan Pada PT. PLN (Persero). *Jurnal Ilmiah Metansi (Manajemen Dan Akuntansi)*, 5(1), 57-62. <https://doi.org/10.57093/metansi.v5i1.157>
8. Sudarshan Kumar. A critical study of financial performance evaluation of Mahindra and Mahindra and Tata motors. *Int J Res Finance Manage* 2019;2(1):68-73. doi: 10.33545/26175754.2019.v2.i1a.124

9. Rahmansyah, A., Yuliawan, D., & Gunarto, T. (2024). The Effect of Electricity Consumption, FDI, and Unemployment on Economic Growth in Indonesia 1990-2021. *Revenue Journal: Management and Entrepreneurship*, 2(1), 53–62. <https://doi.org/10.61650/rjme.v2i1.568>
 10. M. Maaouane, S. Zouggar, G. Krajačić, and H. Zahboune, “Modelling industry energy demand using multiple linear regression analysis based on consumed quantity of goods,” *Energy*, vol. 225, 2021, doi: 10.1016/j.energy.2021.120270
 11. Kim HY. Statistical notes for clinical researchers: simple linear regression 3 - residual analysis. *Restor Dent Endod*. 2019 Feb 12;44(1): e11. doi: 10.5395/rde.2019.44. e11. PMID: 30834233; PMCID: PMC6387894
 12. Chen, C. H., & Hsiao, C. M. (2021). Higher Education Expansion and Credit Risk: Evidence from China Universities Panel Data. *International Journal of Public Administration*, 44(6), 477–488. <https://doi.org/10.1080/01900692.2020.1729183>
 13. Hansen, B. E. (1995). *Time Series Analysis* by James D. Hamilton. Princeton University Press, 1994. *Econometric Theory*, 11(3), 625–630. <https://doi.org/10.1017/S0266466600009440>
 14. Jung, Y., Tian, J., & Bareinboim, E. (2021). Estimating Identifiable Causal Effects through Double Machine Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(13), 12113-12122. <https://doi.org/10.1609/aaai.v35i13.17438>
 15. Bollen, K. A., & Jackman, R. W. (1985). *Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases: An Expository Treatment of Outliers and Influential Cases*. *Sociological Methods & Research*, 13(4), 510-542. <https://doi.org/10.1177/0049124185013004004>
- How to cite this article: Agista Surya Bawana, Nurcahya Yulian Ashar, Humam Rosyadi, Arya Wicaksana. Correlation analysis of the effectiveness of electric energy production at PT PLN (Persero) based on panel data regression method. *International Journal of Research and Review*. 2025; 12(12): 338-345. DOI: <https://doi.org/10.52403/ijrr.20251239>
