Analyzing the Factors Affecting the Electricity Cost in Indonesia: A Multiple Linear Regression Approach

Yanyu Tu*

ADCOTE SCHOOL Shanghai, Shanghai, 201703, China

*sherry0520yu@outlook.com

Abstract. With profound changes in the global economic landscape and the accelerated energy transition, electricity cost management has become a critical issue affecting the sustainable development of national industrial competitiveness and livelihood security. As ASEAN's largest economy and a G20 member, Indonesia's energy consumption structure exhibits typical developing country characteristics: rapid growth in electricity demand driven by economic growth and urbanization, and relatively underdeveloped energy infrastructure and a need to improve energy efficiency. This study, based on 312 Indonesian electricity consumption data obtained from Kaggle, empirically analyzed the key factors influencing electricity costs using multiple linear regression models. The results demonstrate that the constructed model has strong explanatory power, explaining 79.8% of the variation in electricity costs (R^2=0.798). The study found that site area (B=0.606), water consumption (B=0.085), and utilization rate (B=9.254) have a significant positive impact on electricity costs, while factors such as building type and recycling rate have no significant impact. The research results provide empirical evidence for the Indonesian government and enterprises to formulate differentiated electricity price policies and optimize energy management. It is recommended to effectively control electricity costs by controlling site size, improving water use efficiency and managing facility utilization.

Keywords: Electricity Cost, Regression Analysis, Influencing Factors, Indonesia, Energy Management.

1 Introduction

As the largest economy in Southeast Asia, Indonesia has maintained an average annual economic growth rate of 5% in recent years, accompanied by rapid growth in energy consumption. According to the 2023 annual report of the Indonesian Ministry of Energy and Mineral Resources, the country's total electricity consumption increased by 5.8% over the previous year, of which industrial electricity consumption increased by 7.2%, commercial electricity consumption increased by 6.5%, and residential electricity consumption increased by 5.3% [1].

X. Pan et al. (eds.), *Proceedings of the 2026 11th International Conference on Financial Innovation and Economic Development (ICFIED 2026)*, Advances in Economics, Business and Management Research 382,
https://doi.org/10.2991/978-94-6239-642-5_77

However, Indonesia's power infrastructure construction is still relatively lagging behind, and the transmission and distribution loss rate is as high as 9.5%, far higher than the average level of 6.8% in Southeast Asia.

In this context, electricity cost control has become a key issue of concern to the Indonesian government and enterprises. In particular, the intensified global energy price fluctuations since 2022 have further amplified the impact of electricity costs on economic operations. Existing research indicates that factors influencing electricity costs include both hardware factors, such as building characteristics and equipment efficiency, and soft factors, such as operational management and user behavior. However, existing research has largely focused on macro-level analysis or electricity consumption patterns within specific sectors (such as industry or residents). There is a lack of systematic empirical research examining how multiple factors, such as building characteristics, operational efficiency, and environmental practices, jointly influence electricity costs. In particular, for Indonesia, a country with a unique energy structure and development stage, no research has comprehensively utilized cross-sectoral data to develop a comprehensive electricity cost analysis model, quantifying the relative impact and statistical significance of each factor. This research gap limits our understanding of the mechanisms driving electricity costs in Indonesia and undermines the relevance and effectiveness of relevant policies.

This study constructs a comprehensive analytical framework encompassing building physical parameters, operational indicators, and environmental practice variables. Using a multivariate linear regression approach, based on data from 312 cross-sector samples, we quantitatively analyze the impact and statistical significance of each factor on electricity costs. The authors collected electricity usage data from the Kaggle platform, screened valid samples, and used statistical analysis software to develop a multivariate linear regression model to empirically analyze the key factors influencing fluctuations in electricity utilization. It aims to fill the empirical gap in the research on the drivers of electricity costs in Indonesia and provide empirical evidence for policymakers and business managers.

2 Literature Review

The factors affecting electricity costs are a multidisciplinary research field involving energy economics, building science, environmental management, and operations management. Existing studies generally suggest that electricity costs are not determined by a single factor, but are the result of the combined effects of macro policies, micro operations, and physical characteristics.

In terms of the macro electricity price formation mechanism, mathematical methods such as Lagrange multipliers are often used to solve power system optimization problems. A study on the real-time electricity price strategy of smart grids applied the Lagrange multiplier method as the core optimization algorithm of its model [2]. At the micro level, a large number of studies focus on the impact of building characteristics and user behavior on electricity consumption. A systematic review of the driving factors of building operation energy consumption clearly pointed out that the physical scale of the building (such as area) is one of the core influencing factors [3]. At the micro-operation level, the operating efficiency of the equipment is closely related to the electricity cost. Research on constructing a model for predicting electricity connection costs also used the multivariate regression method to analyze the influence of key factors, verifying the effectiveness and practicality of this method in the field of electricity cost prediction [4].

In addition, the impact of operational efficiency and management level on energy consumption cannot be ignored. A study on office buildings in South Jakarta found that implementing operational efficiency improvement measures (such as installing building automation systems, replacing high-efficiency lighting, etc.) can significantly reduce the energy consumption and carbon emissions of buildings [5]. Additionally, those empirical results echoed the findings of another study [6], jointly confirming that by improving water use efficiency and managing water use behavior, household energy consumption expenditure can be effectively affected, which provides a scientific reference for formulating coordinated energy and water conservation management strategies. A community study in Jakarta revealed that environmental awareness represented by recycling behavior may lead to energy efficiency improvements [7]. In terms of research methods, Multiple Linear Regression is a common and effective tool for analyzing the impact of multiple factors on electricity costs. This method was successfully used to construct an MLR-GA-WNN building energy consumption prediction model, which accurately identified key influencing factors and verified the feasibility of the proposed model for public building energy consumption prediction [8].

In summary, existing research provides a solid foundation for understanding the formation mechanism of electricity costs. However, in view of the specific national conditions of Indonesia, most studies are either macro-analyses or targeted at specific types of buildings (such as commercial or residential). There is currently a lack of research that can integrate multidimensional variables such as physical characteristics, operational efficiency, and environmental practices to establish a comprehensive electricity cost analysis model based on Indonesian field data. This study aims to fill this gap and use a multiple linear regression model to systematically evaluate the independent impact and joint explanatory power of various factors on Indonesia's electricity costs, providing a more comprehensive perspective for understanding the country's electricity cost driving mechanism.

3 Methodology

3.1 Source and Description

The data for this study primarily comes from the Kaggle public dataset "Electricity Cost Prediction Dataset," which contains electricity consumption records and related parameters from multiple regions in Indonesia between 2022 and 2023. The dataset covers variables such as site area, building type, average daily water consumption, recycling rate, utilization rate, air quality index, problem resolution time, and number of residents, with a total of 312 valid samples.

3.2 Variable Selection and Processing

Variable selection is based on relevant theoretical support from energy economics and building science. The data used in this paper, such as site area, structure type, water consumption, and recycling rate, were collected from the Kaggle website. In this study, electricity cost is designated as the dependent variable (Y), while the factors hypothesized to influence it are treated as independent variables (X). The variables used in this study and their descriptions are included in Table 1.

Table 1. The Definition of Variables

Symbol	Variable	Description
Y	electricity_cost (USD/month)	The target variable, representing the monthly electricity expenditure, influenced by all other parameters.
X1	Site_area (square meters)	Represents the physical size of the property, a primary driver of energy needs.
X2	Structure_type	Categorizes the property into 'Residential=1', 'Commercial=2', 'Mixed-use=3', or 'Industrial=4', reflecting diverse operational and consumption patterns.This categorical variable was encoded into dummy variables for regression analysis, with 'Residential' as the baseline category.
X3	Water_consumption (liters/day)	An indirect indicator of energy use, particularly for water heating and pumping.
X4	Recycling_rate (%)	Suggests the environmental consciousness and potential energy efficiency practices of the site.
X5	Utilization_rate (%)	Indicates the operational intensity and capacity usage, directly impacting energy demand.
X6	Air_quality_index (AQI)	Can influence energy consumption through HVAC system usage for maintaining indoor air quality.
X7	Issue_resolution_time (hours)	A proxy for operational efficiency; longer resolution times may correlate with energy waste.
X8	Resident_count (number of people)	A significant factor for residential and mixed-use properties, directly affecting electricity usage.

3.3 Introduction of Methods

Multiple linear regression is a statistical method used to analyze the relationship between multiple independent variables (predictors) and a single dependent variable (response). In a multiple linear regression model, the dependent variable is assumed to have a linear relationship with the independent variables. In the real world, a single variable often cannot fully explain complex phenomena, as they are often the result of the interaction of multiple factors. Therefore, using multiple independent variables for joint prediction or estimation can provide more accurate and realistic results. In general, assuming n different predictor variables, the formula of the multiple linear regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \tag{1}$$

Where, Y is the dependent variable; $X_1 \ldots X_n$ is the independent variable; β_0 is the intercept term; $\beta_1 \ldots \beta_n$ is the regression coefficient, which represents the effect of each independent variable on the dependent variable; ε is the error term, which represents the random variation that the model cannot explain.

To enhance the credibility of the model, we conducted a comprehensive model test: in addition to the reported F test and p value, we also calculated the adjusted R^2 (Adjusted R^2) value of 0.791, which is close to the R^2 value (0.798), indicating that the model fits well and is not overfitted. The Durbin-Watson statistic of 1.98 (close to 2) indicates no significant autocorrelation in the residuals. The variance inflation factor (VIF) values are all lower than 3 (see Table 4), indicating that there is no serious multicollinearity problem between the independent variables. Residual analysis shows that the residuals are approximately normally distributed and the variances are homogeneous, meeting the basic assumptions of linear regression.

3.4 Analytical Approach

A hierarchical regression analysis method was used, first incorporating physical characteristic variables (site area, building type), then adding operational efficiency variables (utilization rate, problem resolution time), and finally introducing environmental practice variables (recovery rate). This approach evaluated the incremental contribution of these variables to electricity costs.

4 Results and Discussion

4.1 Model Fit and Validation

The overall model fit is good, as confirmed by the key goodness-of-fit indicators presented in Table 2. The adjusted R^2 value is 0.791, indicating that the independent variables explain 79.1% of the variance in the dependent variable. The Durbin-Watson statistic is 1.98, indicating that there is no autocorrelation between the residuals and that the model specification is reasonable.

Table 2. Regression Model Goodness of Fit Indicators

Indicator	Value	Standardized
R^2	0.798	>0.6 indicates good fit
Adjusted R^2	0.791	>0.6 indicates good fit
Durbin-Watson Statistic	1.98	1.5-2.5 indicates no autocorrelation
F-Statistic	8.91	p=0.000

4.2 Correlation Analysis

Table 3 shows the correlations between the eight variables used in this study: Site_area, Structure_type, Water_consumption, Recycling_rate, Utilization_rate, Air_quality_index, Issue_resolution_time, and Resident_count. The correlation coefficient is used to represent the correlation between electricity cost and these variables. The detailed analysis is as Table 3:

Table 3. Standardized Coefficients (Beta)

Symbol	Variable
Site_area	0.733
Structure_type	-0.072
Water_consumption	0.167
Recycling_rate	-0.074
Utilization_rate	0.180
Air_quality_index	-0.007
Issue_resolution_time	0.033
Resident_count	0.101

The standardized coefficients for structure_type, recycling_rate, and air_quality_index are negative, indicating a negative correlation between these three independent variables and the electricity cost in Indonesia. The remaining variables have positive standardized coefficients, indicating a positive correlation between site_area, water_consumption, utilization_rate, issue_resolution_time, and resident_count.

Table 4. Unstandardized coefficient B and standard error

	Standard Error	B
Constant	284.459	478.672
Site_area	0.061	0.606
Structure_type	52.915	-59.855
Water_consumption	0.036	0.085
Recycling_rate	2.267	-3.390
Utilization_rate	2.498	9.254
Air_quality_index	0.847	-0.124
Issue_resolution_time	2.397	1.592
Resident_count	0.695	1.019

A linear regression analysis was conducted using electricity cost as the dependent variable. As shown in the Table 4, the model R-squared value is 0.798, indicating that site area, structure type, water consumption, recycling rate, utilization rate, air quality index, issue resolution time, and resident count explain 79.8% of the variation in electricity cost.

Electricity_cost = 478.672 + 0.606 * Site_area - 59.855 * Structure_type + 0.085 * Water_consumption - 3.390 * Recycling_rate + 9.254 * Utilization_rate - 0.124 * Air_quality_index + 1.592 * Issue_resolution_time + 1.019 * Resident_count

4.3 Linear Regression

An F-test was performed on the model, and it was found to pass ($F=8.91$, $p=0.000<0.05$). This indicates that electricity cost is influenced by at least one of the following variables: site area, structure type, water consumption, and utilization rate. This partially supports my hypothesis that these factors influence the variation in electricity cost.

Table 5. Result of Linear Regression

Symbol	t	p	VIF	tolerance
Constant	1.683	0.096*	-	-
Site_area	9.852	0.000***	2.486	0.402
Structure_type	-1.131	0.261	1.830	0.546
Water_consumption	2.374	0.020**	2.213	0.452
Recycling_rate	-1.495	0.138	1.095	0.914
Utilization_rate	3.705	0.000***	1.063	0.941
Air_quality_index	-0.146	0.884	1.033	0.968
Issue_resolution_time	0.664	0.508	1.086	0.921
Resident_count	1.465	0.146	2.125	0.471

Finally, as shown in Table 4 and Table 5, the detailed regression output indicates that the p-values for site_area, water_consumption, and utilization_rate are less than 0.05, confirming their statistical significance. The regression coefficient for site_area is 0.606 ($t=9.852$, $p=0.000^*<0.05$). This value is positive and significantly positively correlated with electricity cost. The regression coefficient for utilization_rate is 9.254 ($t=3.705$, $p=0.000^*<0.05$), also significantly positively correlated with electricity cost. The regression coefficient for water_consumption is 0.085 ($t=2.374$, $p=0.020^*<0.05$). Although relatively small, it also demonstrates a significant positive impact.

Notably, the regression coefficients for structure_type are -59.855 and -3.390 for recycling_rate. Although both are negative, indicating a negative correlation with electricity costs, their p-values are both greater than 0.05 (0.261 and 0.138, respectively), indicating that these negative relationships are not statistically significant. The non-significant effect of Structure_type warrants further discussion. This lack of significance could stem from high internal variability in energy-use patterns within each building category, meaning that differences between, for example, individual commercial buildings are as large as those between commercial and residential categories, thereby diluting the overall group effect. Alternatively, the influence of building type might be captured and more directly expressed by other correlated variables in the model, such as Site_area and Resident_count, which serve as more continuous and potent proxies for the underlying energy demand drivers that building type intends to represent. Similarly, the regression coefficient for Air_quality_index is -0.124 (p=0.884), the regression coefficient for Issue_resolution_time is 1.592 (p=0.508), and the regression coefficient for Resident_count is 1.019 (p=0.146). The p-values for these variables are all greater than 0.05, indicating that their impact on electricity cost is not statistically significant. In summary, Site_area, Water_consumption, and Utilization_rate have a significant positive impact on electricity cost. While Structure_type, Recycling_rate, and Air_quality_index show a negative relationship, it is not significant. Issue_resolution_time and Resident_count, while showing a positive relationship, are also not significant.

5 Conclusion

This study, using a multivariate linear regression model, conducted an in-depth analysis of several key factors influencing electricity costs in Indonesia. Empirical results indicate that site area, water consumption, and utilization rate have a significant positive impact on electricity costs, all reaching statistical significance ($p<0.05$). Site area had the most pronounced effect (Beta=0.733, $p<0.001$), indicating that every 100 square meter increase in building area increases monthly electricity costs by approximately US$60.6. The impact of utilization rate was also significant (Beta=0.180, $p<0.001$), with every 1 percentage point increase in utilization rate corresponding to a US$9.254 increase in electricity costs.

In contrast, while building type, recycling rate, and air quality index exhibited negative correlations, these were not statistically significant ($p>0.05$), indicating that the impact of these factors on electricity costs was not fully validated in this study. Similarly, although the coefficients of issue resolution time (Issue_resolution_time) and number of residents (Resident_count) are positive, they are not statistically significant. For policymakers and business managers, these findings suggest that interventions aimed at optimizing building footprints, improving water-use efficiency, and managing operational intensity (utilization rate) could be effective strategies for controlling electricity costs.

However, this study also has some limitations. First, the samples are mainly from urban areas, and rural areas are not representative enough. Additionally, this study is subject to limitations inherent in variable measurement. Certain constructs, such as environmental consciousness and operational efficiency, were proxied by indirect indicators (e.g., Recycling_rate and Issue_resolution_time). While practical, these proxies may not fully or precisely capture the underlying, complex behaviors and managerial practices they are intended to represent. This potential measurement error could attenuate the observed relationships, potentially explaining the lack of statistical significance for some variables in the model. Moreover, the impact of external factors such as regional electricity price policies, seasonal changes and climate conditions is not fully considered.

Future research can conduct in-depth exploration in the following aspects: first, expand the scope and number of samples, especially increase the samples of industrial users and rural areas [9]; second, use panel data models to analyze the long-term trend of electricity costs; third, introduce more influencing factors, such as the proportion of renewable energy use [10], smart meter coverage, etc.; fourth, try to use advanced algorithms such as machine learning to improve the accuracy of electricity cost prediction. The theoretical contribution of this study is to construct a comprehensive analysis framework that integrates physical characteristics, operational efficiency and environmental practices, and confirms the important role of operational efficiency factors in explaining electricity cost variations. In practice, it provides an evidence-based decision-making basis for Indonesia's electricity cost control, which helps promote the country's energy efficiency and sustainable development.

References

1. Institute for Essential Services Reform.: Indonesia Energy Transition Outlook 2023. IESR, Jakarta (2023).
2. Wang, W., Li, J., Zhang, C.: Research on real-time electricity pricing strategy for smart grid based on Lagrange multipliers. Automation of Electric Power Systems 47(15), 112–121 (2023).
3. Ye, H., Yan, H., Zhang, R., Zhao, Z., Lin, T., Zhang, G.: Research progress on driving mechanism of building operation energy consumption for low-carbon city construction. Chinese Journal of Applied Ecology 32(7), 2644–2652 (2021).
4. Wang, H., Sun, Z.: Connection cost prediction model based on power big data analysis. Power Demand Side Management 25(5), 104–109 (2023).
5. Farizal, F., Noviandri, M.A., Hamdani, H.: Sustainability Development through a Nearly Zero Energy Building Implementation Case: An Office Building in South Jakarta. Sustainability 16(16), 1–20 (2024).
6. Wang, C.Y., Li, Z.H., Ni, X.Y., et al.: Residential water and energy consumption prediction at hourly resolution based on a hybrid machine learning approach. Water Research 246, 120–135 (2023).
7. Chen, Y., Abdul, R.: Energy Efficiency and Recycling Practices in Jakarta: A Community-Based Study. Journal of Environmental Management 351, 119876 (2024).
8. Ye, Y., Ma, H., Li, S.: Research on energy consumption prediction of public buildings based on MLR-GA-WNN. Computer Simulation 40(3), 514–519 (2023).
9. World Bank.: Indonesia Economic Quarterly: Investing in Tomorrow. World Bank Group, Washington, DC (2023).
10. International Energy Agency.: Energy Efficiency 2023. IEA Publications, Paris (2023).