International Journal of Innovation Scientific Research and Review
Vol. 04, Issue, 05, pp.2693-2695, May 2022
Available online at http://www.journalijisr.com
SJIF Impact Factor 4.95

ISSN: 2582-6131

Research Article

APPLICATION OF MULTIPLE LINEAR REGRESSION ANALYSIS TO FORECAST ELECTRICITY CONSUMPTION IN CHONBURI PROVINCE

Jatupat Mekparyup and * Kidakan Saithanu

Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

Received 05th March 2022; Accepted 06th April 2022; Published online 20th May 2022

ABSTRACT

This research studied forecasting electricity consumption in Chonburi province, Thailand. Multiple linear regression method was applied in this research using three crucial variables; number of electricity customers, rainfall and temperature to fit the equation. All data was separated by two data sets, training and validation. The first set was used to generate the predicting equation and the other was to validate the accuracy performance with mean absolute percentage error (MAPE). All assumptions of regression analysis were tested by Anderson-Darling statistic, Durbin-Watson statistic, Breusch-Pagan statistic and Variance inflation factor. Finally, the performance of predicting was monitored. The results displayed the regression standard error, the adjusted coefficients of determination and MAPE with 2,299,595, 0.879 and 3.853%, respectively.

Keywords: electricity consumption, regression analysis.

INTRODUCTION

Electricity consumption rate of the Eastern region of Thailand increases 7.22 percent in the first half of 2016 compared to the same period in the last year. Mainly one factor is the expansion of production group by 11.37 percent which is an indication of the growth factor in the manufacturing industry under the Eastern Economic Corridor (EEC) (Energy Regulatory Commission, 2016, p. 10). Chonburi is the one of three provinces set in the special zone of EEC that has a large international shipping port, Laem Chabang Port. In addition, Chonburi province has many large industrial estates divided by four districts; Muang Chonburi, Sri Racha, Ban Bueng and Phan Thong. For this reason, Chonburi province tends to increase electricity consumption in the future clearly illustrated in Fig 1. Then the main purpose of this research is to forecast electricity consumption in Chonburi province with multiple regression analysis.

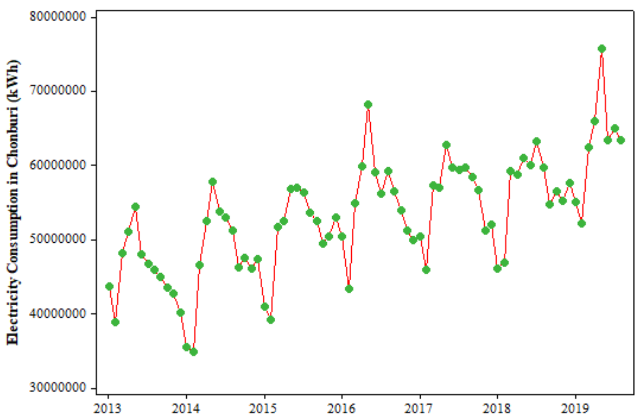

Figure 1: Electricity consumption in Chonburi province since January 2013 – August 2019

***Corresponding Author: Kidakan Saithanu,**
Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

MATERIALS AND METHODS

The three significant factors affecting to electricity consumption; number of electricity customers, rainfall and temperature, were provided by District 2 Central Electricity Authority in Chonburi since January 2013 to August 2019 divided into two data sets; training and validation data set shown in Table 1.

Table 1: Data set

Data set	Size	%
Training (January 2013 – December 2017)	60	75
Validation (January 2018 – August 2019)	20	25
Total	80	100

The multiple regression model consists of one dependent variable, electricity consumption (y; kWh), and three independent variables; number of electricity customers (x_1), rainfall (x_2; millimeters) and temperature (x_3; Celsius), as of equation (1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \qquad (1)$$

where β_0 is y-intercept of the regression model, β_1, β_2, β_3 are regression coefficients and ε represents for an error term. Correlation coefficient was firstly screened to depict linear relationship between dependent variable and the three independent variables by Pearson correlation. The multiple linear regression equation was then fitted by stepwise method. After fitting multiple regression equation, the assumptions of regression analysis were validated with Anderson-Darling statistic (AD)to detect normality (Anderson and Darling, 1952), Durbin-Watson statistic (DW) to catch auto-correlation (Durbin and Watson, 1950), Breusch-Pagan statistic (BP) to validate homoscedasticity (Breusch and Pagan, 1979) and Variance inflation factor (VIF) to monitor multicollinearity respectively. Finally, the performance of forecasting was computed with mean absolute percentage error (MAPE).

RESULTS

Descriptive statistics were measured by mean, standard deviation (SD), minimum value and maximum value demonstrated in Table 2.

Table 2: Data summary

Monthly average data	Mean	SD	Min	Max
Electricity consumption (y)	53,222,105	7,605,298	35,003,594	75,803,681
Number of electricity customers (x_1)	159,033	16,939	128,089	182,975
Rainfall (x_2)	10.328	5.786	1.570	27.100
Temperature (x_3)	28.305	1.253	24.560	30.790

Graphical linear relationship among 4 variables was created by scatter plot in Fig 2. Also, the numerical one among Electricity consumption (y), number of electricity customers (x_1), rainfall (x_2) and temperature (x_3) were generated by Pearson correlation coefficients as shown in Table 3. The highest positive value was significantly detected between y and x_1 with 0.593 (p-value<0.000).

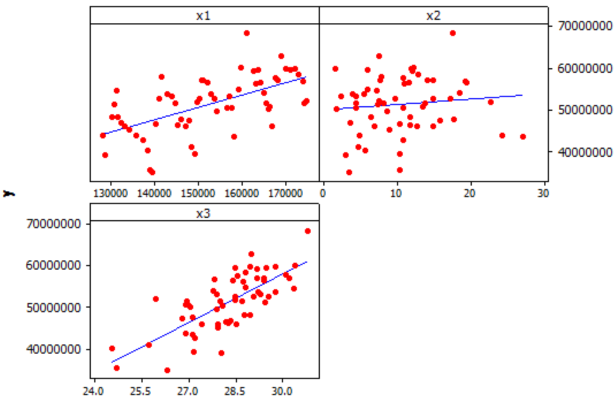

Figure 2: Scatter plot between dependent and three independent variables of training data set

Table 3: Pearson correlation coefficients among four variables

Variables	Variables			
	y	x_1	x_2	x_3
y	1.000 (0.000)	- -	- -	- -
x_1	0.593 (0.000)	1.000 (0.000)	- -	- -
x_2	0.106 (0.422)	0.235 (0.071)	1.000 (0.000)	- -
x_3	0.746 (0.000)	0.058 (0.659)	-0.029 (0.827)	1.000 (0.000)

p-values in bracket

The linear equation was built by multiple regression analysis with stepwise method as equation (2).

$$\hat{y} = -101,326,393 + 279.36x_1 + 3,895,888x_3 \qquad (2)$$

After fitting the equation, the multiple linear regression model was tested by analysis of variance (ANOVA) in Table 4.

Table 4: ANOVA

Source of variation	Degree of freedom	Sum of square	Mean of square	F	p-value
Regression	2	2.198×10^{15}	1.099×10^{15}	207.86	0.000
Residual	52	2.908×10^{14}	5.288×10^{12}		
Total	57	2.489×10^{15}			

It was illustrated that the model was respectively appropriated with *F*-statistic values 207.86 (p-value = 0.000) with standard error of regression (*S*) of 2,299,595 and the adjusted coefficients of determination (r^2_{adj}) of 0.879. All four regression assumptions; normality, auto-correlation, homoscedasticity and multicollinearity, were validated in Table 5. Normality testing by AD statistic value was 0.871 (p-value = 0.128), auto-correlation testing by DW statistic value was 2.548 with upper critical value D_U = 1.645, homoscedasticity testing by BP statistic value was 3.625 (p-value = 0.837) and multicollinearity testing by VIF was 1.0 with both x_1 and x_2. The results lastly found that testing all assumptions was satisfied.

Table 5: Assumption testing

Assumptions	Test statistic	Critical value	p-value
Normality	AD = 0.571	-	0.128
Auto-correlation	DW = 2.548	D_U = 1.645	-
Homoscedasticity	BP = 3.625	5.991	0.837
Multicollinearity	VIF (x_1) = 1.0	-	-
	VIF (x_3) = 1.0	-	-

Finally, the performance of forecasting was presented with time series between real and predicted electricity consumption were plotted in Fig 3 and numerical performance index was calculated by MAPE with 3.853%.

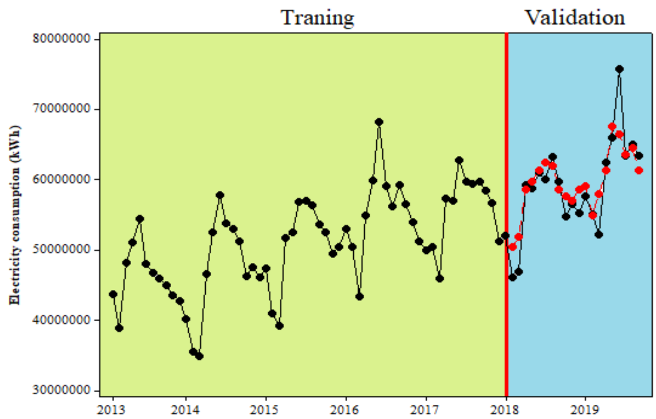

Figure 3: Time series plot between real and predicted data of electricity consumption

CONCLUSION AND DISCUSSION

Multiple regression equation was built to predict the average of monthly electricity consumption in Chonburi province. The two significant factors effecting to the electricity consumption were number of electricity customers and temperature. The standard error of regression was 2,299,595 and the adjusted coefficients of determination was 0.879with mean absolute percentage error of 3.853%.

ACKNOWLEDGEMENTS

Specially, the authors were appreciated to provide the secondary data from District 2 Central Electricity Authority in Chonburi.

REFERENCES

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. The annals of mathematical statistics, 193-212.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the econometric society, 1287-1294. Durbin,

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. Biometrika, 37(3/4), 409-428.

Energy Regulatory Commission. (2016). Overview of electricity consumption by economic sector. Available: https://www.erc.or.th/ERCWeb2/default.aspx

Koivisto, M., Heine, P., Mellin, I., & Lehtonen, M. (2012). Clustering of connection points and load modeling in distribution systems. IEEE Transactions on Power Systems, 28(2), 1255-1265.

Lam, J. C., Wan, K. K., Liu, D., & Tsang, C. L. (2010). Multiple regression models for energy use in air-conditioned office buildings in different climates. Energy Conversion and Management, 51(12), 2692-2697.

Montgomery, D. C., & Askin, R. G. (1981). Problems of nonnormality and multicollinearity for forecasting methods based on least squares. AIIE Transactions, 13(2), 102-115.

Sanquist, T. F., Orr, H., Shui, B., & Bittner, A. C. (2012). Lifestyle factors in US residential electricity consumption. Energy Policy, 42, 354-364.

Varisara, T. (2017). Inverse modeling for energy consumption prediction in higher educational
buildings: a case study of Chulalongkorn University. Master's thesis, Department of Architecture, Faculty of Architecture, Chulalongkorn University

****