International Journal of Innovation Scientific Research and Review
Vol. 05, Issue, 01, pp.3766-3770, January 2023
Available online at http://www.journalijisr.com
SJIF Impact Factor 4.95

ISSN: 2582-6131

Research Article

ELECTRICITY CONSUMPTION FORECASTING BASED ON MULTIPLE REGRESSION ANALYSIS

Jatupat Mekparyup and *Kidakan Saithanu

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

Received 06th November 2022; Accepted 07th December 2022; Published online 22th January 2023

ABSTRACT

This study was to forecast monthly average of electricity consumption, the response variable, in Trat province, Thailand with stepwise method of multiple linear regression technique using the three particularly explanatory variables; monthly average number of electricity customers, monthly average rainfall and monthly average temperature. Data was divided into training and validation data sets. The first one was used to build the forecast equation then all assumptions of multiple linear regression analysis were testified by Anderson-Darling statistic (AD), Durbin-Watson statistic (DW), Breusch-Pagan statistic (BP) and Variance inflation factor (VIF). The second one was used to validate the performance of prediction with adjusted coefficients of determination, standard error of regression, mean absolute percentage error, mean absolute error and root mean squared error. The results showed that the number of electricity customers and temperature were significant factors to forecast electricity consumption in Trat province. The performance of prediction accuracy successfully displayed the adjusted coefficients of determination of 0.730, standard error of regression of 93,325.10 kWh, mean absolute percentage error of 4.08%, mean absolute error of 92,116.8812 kWh and root mean squared error of 131,836.3008 kWh.

Keywords: electricity consumption, multiple regression analysis.

INTRODUCTION

Tratis one of the eastern provinces in Thailand with a total area of 2,819.0 sq km. concluding7districts;Muang, Khlong Yai, Khao Saming, Bo Rai, Laem Ngop, Ko Kut and Ko Chang (Provincial Electricity Authority of Trat province, 2022).

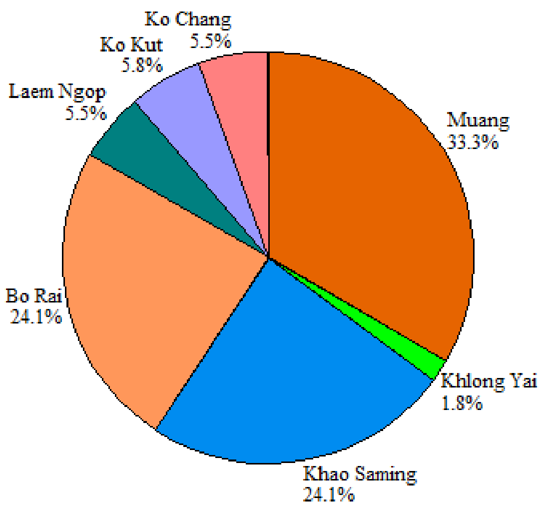

Figure 1: Area of each district in Trat province

According to the electricity consumption statistics in Trat, it was found that the amount of electricity consumption had an overview tendency to yearly increase since 2013 to 2019 especially in March, April and May illustrated in Fig 2. High electricity consumption was found in 3 districts; Khlong Yai, Laem Ngop and Ko Chang, due to a considerable tourist attraction which consists of a large number of hotels. For this reason, the researcher is interested to study the forecast of electricity consumption in Trat province by regression techniques (Bianco, Manca, and Nardini, 2013; Panklib, Prakasvudhisarn, and Khummongkol, 2015; Ezenugu, Nwokonko, and Markson, 2017; Akdi, Gölveren, and Okkaoğlu, 2020).

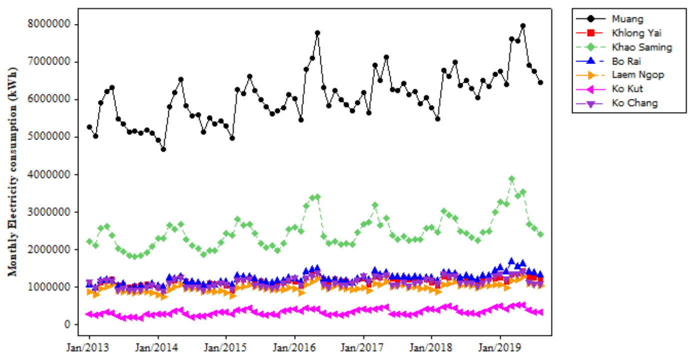

Figure 2: Monthly electricity consumption of all districts in Trat province since January 2013 – August 2019

MATERIALS AND METHODS

The regression technique is one of the most popular methods of forecasting values especially when only one explanatory variable (x) has a linear relationship with the response variable (y) known simple linear regression instead of two or more explanatory variables are called multiple linear regression.The three explanatory variables; monthly average number of electricity customers, monthly average rainfall and monthly average temperature, and one response variable, monthly average of electricity consumption, were succeeding applied in this research (Bianco, Manca, and Nardini, 2013; Panklib, Prakasvudhisarn, and Khummongkol, 2015; Ezenugu, Nwokonko, and Markson, 2017; Akdi, Gölveren, and Okkaoğlu, 2020). All data

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

was collected by District 2 Central Electricity Authority in Chonburisince January 2013 to August 2019 then separated into two data sets; training data set (January 2013 to December 2017) and validation data set (January 2018 to August 2019) described 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**

Model fitting

Firstly, linear relationship was examined by Pearson correlation coefficients (r). Multiple linear regression model was then generated by stepwise method to forecast monthly average of electricity consumption in Trat province following by Equation (1)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (2)$$

where y is the monthly average of electricity consumption (kWh), x_1 is the monthly average number of electricity customers, x_2 is the monthly average rainfall (millimeters), x_3 is the monthly average temperature (Celsius), β_0 is y-intercept of the model, β_1, β_2, β_3 are regression coefficients and εrepresents for an error term.

Model testing

After fitted model, an appropriate model was validated by F-test then regression coefficients were also tested by t-test.

Assumptions testing

Normality test was firstly detected by Anderson-Darling statistic (AD) (Anderson and Darling, 1952). Auto-correlation, homoscedasticity (constant variance) and multicollinearity were later monitored by Durbin-Watson statistic (DW) (Durbin and Watson, 1950), Breusch-Pagan statistic (BP) (Breusch and Pagan, 1979) and variance inflation factor (VIF) respectively.

$$\text{AD} = -n - \sum_{i=1}^{n} \left(\frac{2i-1}{n} \right) \{ \ln F(y_i) + \ln[1 - F(y_{n+1-i})] \} \quad (2)$$

$$\text{DW} = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad (3)$$

$$\text{BP} = \frac{SSR^* / 2}{(SSE / n)^2} \sim \chi^2_{(k)} \quad (4)$$

$$\text{VIF}_j = C_{jj} = \frac{1}{1 - R^2_{j|\text{others}}}; \; j = 1,\ 2,\ \ldots,\ k \quad (3)$$

Accuracy validation

The accuracy of prediction was finally validated by adjusted coefficients of determination (r^2_{adj}) and standard error of regression (S). Moreover, mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean squared error (RMSE) were also commonly verified.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \quad (8)$$

RESULTS

Each of three data sets; all, training and validation, was displayed with the values of average (Mean), standard deviation (SD), minimum (Min) and maximum (Max) severally categorized by the monthly average of electricity consumption (y), the monthly average number of electricity customers (x_1), the monthly average rainfall (x_2) and the monthly average temperature (x_3) in Table 2.

Table 2: Data summary

Variables	Data set	Mean	SD	Min	Max
y	All	1,924,017	226,017	1,555,630	2,550,812
	Training	1,859,361	194,480	1,555,630	2,456,440
	Validation	2,117,985	205,299	1,802,694	2,550,812
x_1	All	10,116	472	9,233	10,851
	Training	9,923.1	378.5	9,233	10,521
	Validation	10,695	97	10,539	10,851
x_2	All	19.990	13.310	0.000	53.180
	Training	20.270	13.500	0.000	53.180
	Validation	19.130	13.010	0.467	43.550
x_3	All	27.748	0.771	25.594	29.773
	Training	27.742	0.798	25.594	29.773
	Validation	27.766	0.704	26.771	29.320

Model fitting

Linear relationship between the response variable (y) and three the explanatory variables (x_1, x_2 and x_3) was explored with graphical display as Figure (3).

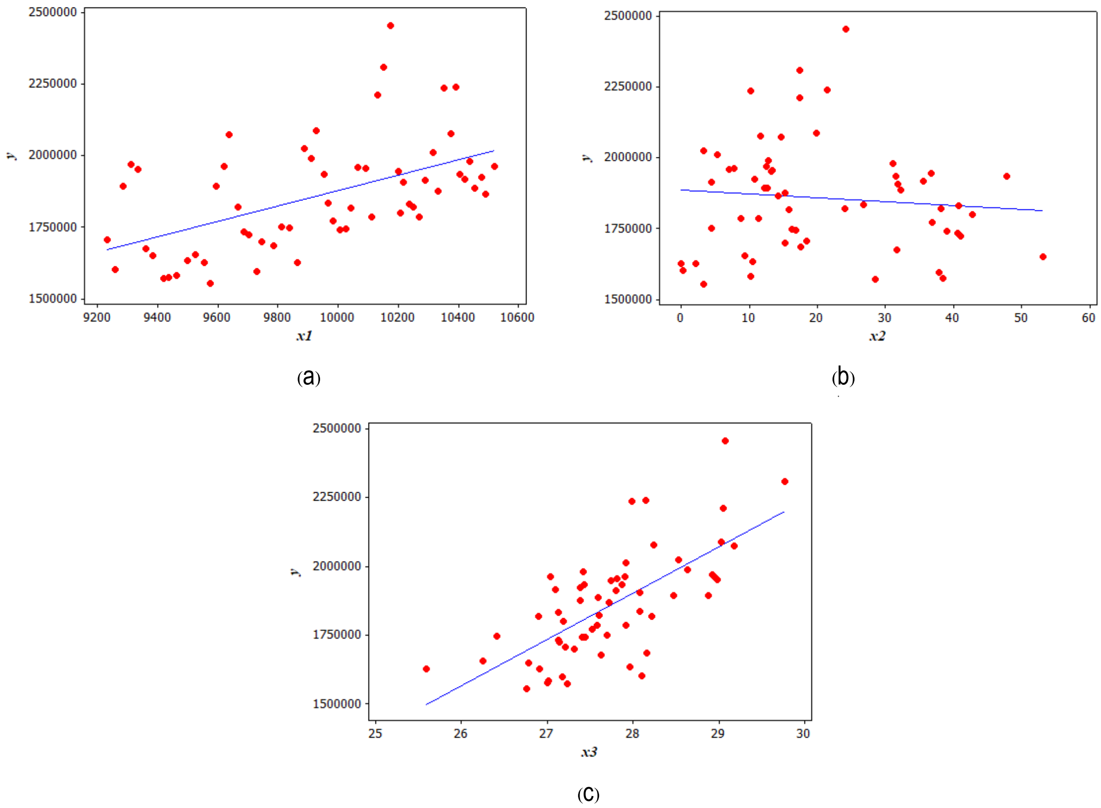

Figure 3: Scatter plot between (a) y and x_1, (b) y and x_2, (c) y and x_3

As of Figure (3), the relationship between y and x_1 and y and x_3 were quite linear. It was confirmed by Pearson correlation coefficients that $r_{y,x1}$ was 0.524 (p-value = 0.000) and $r_{y,x3}$ was 0.690 (p-value = 0.000) displayed in Table 3.

Table 3: Pearson correlation coefficients among four variables

Variables	Variables		
	y	***x*****1**	***x*****2**
x_1	**0.524**		
	(0.000)		
x_2	-0.093	0.055	
	(0.481)	(0.676)	
x_3	**0.690**	0.023	-0.186
	(0.000)	(0.863)	(0.155)

p-values in bracket

The step wise method was employed to generate multiple linear regression equation for forecasting the monthly average of electricity consumption (y) as of following Equation (9).

$$\hat{y} = -4{,}909{,}020 + 251.92x_1 + 153{,}663x_3 \quad (9)$$

Model testing

After fitted the equation, adequacy of the model was checked by F statistic value of 79.29 (p-value = 0.000). It was testified that the model was appropriate for forecasting.

Table 4: ANOVA

Source of variation	Degree of freedom	Sum of square	Mean of square	*F*	p-value
Regression	2	1.3812×10^{12}	6.9062×10^{11}	79.29	0.000
Residual	56	4.8774×10^{11}	87,095,666,620		
Total	58	1.8690×10^{12}			

Consequently, two parameters of regression coefficients were examined by *t* statistic value of 7.82 (p-value = 0.000) and 9.84 (p-value = 0.000). It was claimed that each parameter was not equal to zero therefore this model was suitable for forecasting.

Table 5: Regression coefficients

Estimator	Coefficient values	*t*-value	*p*-value
$\hat{\beta}_0$	-4,909,020	-9.14	0.000
$\hat{\beta}_1$	251.92	7.82	0.000
$\hat{\beta}_3$	153,663	9.84	0.000

Assumptions testing

All regression analysis assumptions were tested by AD = 0.526 (p-value = 0.173), DW = 1.8781 (upper critical value = 1.6497), BP = 0.281 (p-value = 1.31) and all VIFs = 1.00. According to the results as Table 6, it was manifested that all assumptions met the criteria.

Table 6: Assumption testing

Assumptions	Test statistic	Critical value	p-value
Normality	AD = 0.526		0.173
Auto-correlation	DW = 1.8781	D_U = 1.6497	
Homoscedasticity	BP = 0.281	5.991	0.131
Multicollinearity	VIF (x_1) = 1.00		
	VIF (x_3) = 1.00		

Accuracy validation

The graphical accuracy validation was illustrated in Figure 4 and the numerical accuracy validation was calculate by r^2_{adj} = 0.730, S = 93,325.10, MAPE = 4.08%, MAE = 92,116.8812 and RMSE = 131,836.3008.

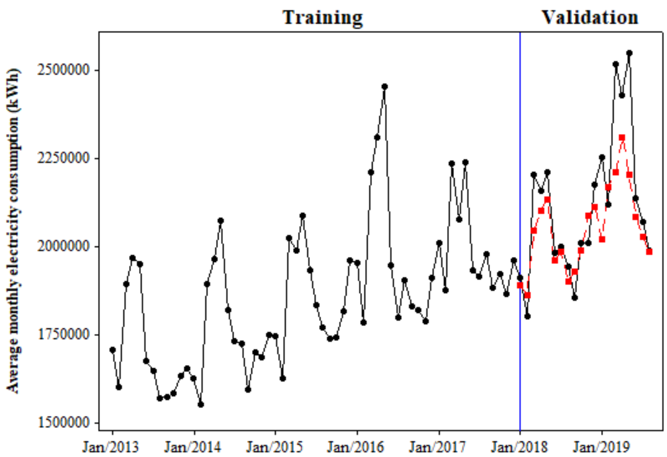

Figure 4: Time series plot between real and predicted data of monthly average of electricity consumption in Trat province

CONCLUSION AND DISCUSSION

The forecasting equation of the monthly average of electricity consumption in Trat province was established by multiple liner regression analysis with stepwise method. The factor had an effect on the electricity consumption was the number of electricity customers and temperature. The index of prediction accuracy was respectively determined by the adjusted coefficients of determination of 0.730, the standard error of regression of 93,325.10 kWh, mean absolute percentage error of 4.08%, mean absolute error of 92,116.8812 kWh and root mean squared error of 131,836.3008 kWh.

ACKNOWLEDGEMENTS

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

REFERENCES

Akdi, Y., Gölveren, E., & Okkaoğlu, Y. (2020). Daily electrical energy consumption: Periodicity, harmonic regression method and forecasting. Energy, 191, 116524.

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.

Bianco, V., Manca, O., & Nardini, S. (2013). Linear regression models to forecast electricity consumption in Italy. Energy Sources, Part B: Economics, Planning, and Policy, 8(1), 86-93.

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

Ezenugu, I. A., Nwokonko, S. C., & Markson, I. (2017). Modelling and Forecasting of residential electricity consumption in Nigeria using Multiple and Quadratic regression models. American Journal of Software Engineering and Applications, 6(3), 99 104.

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.

Panklib, K., Prakasvudhisarn, C., & Khummongkol, D. (2015). Electricity consumption forecasting in Thailand using an artificial neural network and multiple linear regression. Energy Sources, Part B: Economics, Planning, and Policy, 10(4), 427-434.

Provincial Electricity Authority of Trat province. (2022). electricity consumption statistics in Trat. Available: https://www.pea.co.th/

Sanquist, T. F., Orr, H., Shui, B., & Bittner, A. C. (2012). Lifestyle factors in US residential electricity consumption. Energy Policy, 42, 354364.

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

****