Forecasting Electricity Load Demand- An Power System Planning

Elektrotechnik Berg
University of Cologne, Albertus-Magnus-Platz, 50923 Koln, Germany
berg3433@hotmail.com

Article Info
Journal of Journal of Enterprise and Business Intelligence (http://anapub.co.ke/journals/jebi/jebi.html)
Doi: https://doi.org/10.53759/5181/JEBI202101022
Received 25 April 2021; Revised form 25 July 2021; Accepted 25 August 2021.
Available online 05 October 2021.

Abstract- Moving holiday electricity load demand forecasting is one of the most challenging topics in the forecasting area. Forecasting electricity load demand is essential because it involves projecting the peak demand level. Overestimation of future loads results in excess supply. Wastage of this load is not welcome by the international energy network. An underestimation of load leads to failure in providing adequate reserve, implying high costs. Many factors can influence the electricity load demand, such as previous load demand, type of the day, coincidence with other holidays and the impact of major events. Hence, 12 independent variables were considered in constructing the regression model to forecast moving holiday electricity load demand. This study investigates Malaysia's daily electricity load demand data using multiple linear regression to forecast electricity load demand on moving holidays, such as Hari Raya AidilFitri, Chinese New Year, Hari Raya AidilAdha, and Deepavali from September 2016 to October 2017. The result shows six independent variables are significant from the several method variables selections. Overall, the constructed models from this study give promising results and can forecast for next year's moving holiday electricity load demand with a sample forecasting error of 3.7% on the day of the moving holiday.

Keywords - Electricity Load Demand; Linear Regression; Moving Holiday; Time Series Forecasting.

I. INTRODUCTION

Forecasting on electricity load demand (ELD) is essential to support the system and operation of the electric utility business in the future. In addition, it involves projections of peak demand levels and overall energy consumption patterns in electrical loads and demand forecasts. ELD forecasting is carried out to represent the main task in planning electricity production because the source needs to be determined, especially in operating the power plant, such as daily fuel consumption. According to [1], ELD forecasting is a central and integral process for planning periodical operations and facility expansion in the electricity sector. Therefore, the study on ELD forecasting is significant to assist utility companies in developing the power system's efficient operation to balance between generation and load demand [2], [3]. Moreover, this may reduce the power system operational cost [4]. Hence, the modeling of ELD with the minimum forecast error becomes very important to obtain optimum cost and maximize profit. Consequently, many previous studies have conducted forecasting in the ELD area, for example [5].

However, the ELD forecasting error is affected by the moving holiday effect. In fact, a 20% increment of forecasting error has been noted on moving holidays when compared with a normal day [6]. If the moving holiday falls on a Saturday or Monday instead of other weekdays, then there is a chance for the occurrence of a significant load forecasting error [7]. Moreover, according to [8], one percent (1%) reduction of mean absolute percentage error (MAPE) in load forecasting saves 10,000 MW in electricity load, which may lead to savings of approximately £1.6 million (around RM9 million at current exchange rates) per year. This means that a 1% reduction in MAPE in 2020 essentially results in savings of more than £1.6 million per year since the price of electricity has increased annually in real terms.

Therefore, this study focuses on forecasting ELD in Malaysia concerning the relatively unique consumption pattern due to the multi-festival holidays. Malaysia has a diverse ethnic, where most of the population are Malays, followed by Chinese, Indian, etc. Each race has a variety of festivals and festive holidays. The main festivals in Malaysia are usually related to religious activities involving Muslims, Chinese, Hindus and others. In addition, the dates of many festivals are determined based on the lunar calendar.

According to three different calendars, the date of moving holidays is based on the Geogerion calendar; Chinese, Hindu and Hijriah lunar calendar. Therefore, this holiday date does not occur on a fixed date each year but shifts from one period to another for many years. Since the ELD patterns on holidays are often idiosyncratic in nature, this leads to significant predictive errors [8]. Therefore, irregular holidays like Hari Raya AidilFitri, Hari Raya AidilAdha, Chinese New Year and Deepavali from one year to the next may influence the results of predicting time series data. In addition, some of these festivals holidays overlap with other holidays and increase the difficulty of the activity for predicting ELD. Another study has been conducted to understand electricity demand during a special day or holiday.

Numerous studies in different science branches have been conducted to study the effects of relationships between calendar variations. Indeed, the effects of the calendar have been studied in the stock market [6], [9], [10], water demand [29], economic, retailers' warehouses [3] insurance [5] , air pollution [10], [29], social, psychiatric, medical [24] transportation [4], [11], zoological, tourism finance, logistic [12].

Researchers in the field of ELD involving moving holidays are, for example [13]. Research by [14] considered moving holiday effects and, therefore, gave a better forecasting accuracy for Malaysia's peak daily load. Meanwhile, [15] used dynamic regression intervention modeling for the Malaysian daily load based on moving holidays data. On the other hand, Kim, [8] introduced special day as a dummy variable in forecasting ELD models, forecast moving holiday ELD week based on fuzzy time series using a specific weighted mechanism.

Moving holidays should be considered in the seasonal adjustment to avoid misleading interpretations of seasonally adjusted and trend estimates. The study's reason is to forecast the ELD during moving holidays, such as Hari Raya, Deepavali, Chinese New Year, etc. We can determine how much demand for power people need during the holiday and prevent a shortage of power during that holiday. Moreover, the error of ELD forecasting will increase the operational costs. Overestimation of future load results in surplus supply, which is not welcomed by the international energy network.

On the other hand, underestimation of load causes failure in providing adequate reserve and implies high costs. Therefore, this study aims to identify the significant variable that affects the ELD in the model. We then construct a multiple linear regression model for moving holiday ELD and finally, to forecast the ELD three days before the holiday, on the holiday and next, there days after holiday for moving big holiday events (Hari Raya AidilFitri, Chinese New Year) and not big events (Hari Raya AidilAdha, Deepavali). This study focuses on data for a week moving holidays of Hari Raya AidilFitri, Chinese New Year, Deepavali and Hari Raya AidilAdha which is the moving holiday is on the fourth day of the week [16].

II. METHODOLOGY

Data Collection and Variables

For data collection, secondary data have been used collected from Tenaga Nasional Berhad (TNB) on Grid System Operator's website. This study's scope only concentrates on daily ELD for weeks that only have a moving holiday from 1st September 2016 to 31st October 2017 recorded in Malaysia. The data consists of daily data that partition into two parts. Data from 1st September 2016 until 30th September 2017 (35 data) were used to formulate a prediction model, and October 2017 (7 data) to validate the prediction model [17].

This study used electricity load demand, Y as the dependent variable measured using kilowatt (KW) scale. A total of 12 variables were used as the independent variables, as listed in Table 1. To capture the type of the day effect, qualitative of the day has been introduced into the model through the specification of a dummy variable (Z_1) representing the type of that day, which is not a holiday, where the holiday is the base. That is;

$$Z_1 = \begin{cases} 1 & \text{if not holiday} \\ 0 & \text{Otherwise} \end{cases}.$$

To capture the coincidence effect, qualitative of the coincidence has been introduced into the model through the specification of three dummy variables representing coincidence in the day, which is before coincidence (Z_2), after coincidence (Z_3) and no coincidence (Z_4), with the coincidence as the base [18].

For before coincidence;

$$Z_2 = \begin{cases} 1 & \text{if before coincidence} \\ 0 & \text{Otherwise} \end{cases}$$

For after coincidence;

$$Z_3 = \begin{cases} 1 & \text{if after coincidence} \\ 0 & \text{Otherwise} \end{cases}$$

For no coincidence;

$$Z_4 = \begin{cases} 1 & \text{if no coincidence} \\ 0 & \text{Otherwise} \end{cases}.$$

To capture the big event effect, qualitative of the day has been introduced into the model through the specification of a dummy variable (Z_5) representing the type of that week, which is a big event in the week, where no big event's week as the base, given by [19];

$$Z_5 = \begin{cases} 1 & \text{if big event} \\ 0 & \text{Otherwise} \end{cases}$$

Table 1. Dependent and independent variables for ELD

Symbol	Variables
Y	ELD
X_1	The first day of ELD before the holiday
X_2	The second day of ELD before the holiday
X_3	The third day of ELD before the holiday
X_4	The fourth day of ELD before the holiday
X_5	The fifth day of ELD before the holiday
X_6	The sixth day of ELD before the holiday
X_7	The seventh day of ELD before the holiday
Z_1	Dummy variable not a holiday
Z_2	Dummy variable before coincidence
Z_3	Dummy variable after coincidence
Z_4	Dummy variable no coincidence
Z_5	Dummy variable big event

Method of Data Analysis

The objective of this research work is to predict moving holiday ELD using multiple linear regression methods. The developed models are used for predictions of out of sample forecasts. The estimated models can be written [20] as:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 Z_1 + \hat{\beta}_9 Z_2 + \hat{\beta}_{10} Z_3 + \hat{\beta}_{11} Z_4 + \hat{\beta}_{12} Z_5.$$

Since there were so many independent variables in this study, we used the variable selection process to construct the best model that predicts well or explains the data's relationships. There are several selection process methods, which are stepwise selection, backward elimination, and forward selection.

A few assumptions have to be tested in the multiple linear regression analyses because the result is invalid if the assumptions are not met. These assumptions include;

- For any specific value of any of the independent variables, the dependent variable's values are normally distributed.
- There is a linear relationship between the dependent variable and each of the independent variables.
- The observations of the dependent variable are independent of each other.
- The variance for the normal distribution of possible values for the dependent variable is the same for each independent variable's value.

In summary, the assumptions describe the probability distributions of the random error in the model where,

$$\text{Random error} = e = Y - E(Y) = Y - b_0 - b_1 X_1 - \ldots - b_k X_k.$$

III. ERROR MEASUREMENT

Model forecasting performance usually will be compared by using a variety error measurement. The forecasting error in time period t can be defined as the actual value minus the prediction value, $e_t = y_t - \hat{y}_t$, where y_t is the actual value at time t and $\hat{y}_t$ is the fitted value at time t. This study will employ the Mean Square Error (MSE), Absolute Persentage Error (APE) and Mean Absolute Percentage Error (MAPE) to calculate the error measurement. MSE is considered the most accurate measure to define which models avoid large errors because it can discover large forecast errors. The MSE is given as,

$$\text{MSE} = \frac{\sum_t^n e_t^2}{n}.$$

On the other hand, MAPE is used to perform comparisons relative or percentage error measures. To compute MAPE, we must first compute the APE for each forecast. MAPE is computed as follows:

$$\text{MAPE} = \sum_{t=1}^{n} \frac{\left|\left(\frac{e_t}{y_t}\right)\times 100\right|}{n},$$

where n is the number of observation and $\left|\left(\frac{e_t}{y_t}\right)\times 100\right|$ is APE that calculates the fitted values for a particular forecasting method.

Framework of the study

Figure 1 shows the methodology framework of this study. The methodology starts with pre-processing data. In this stage, the collection and cleaning data processes were conducted. Then, in the next stage, the data were partitioned into two parts: estimation and evaluation. The methodology stage continues with the variable selection process conducted in the estimation part of the data using three methods: Stepwise, Backward and Forward variable selection. Next, the regression analysis of assumption checking is performed. After all the assumption tests were satisfied, the next process constructs the significance regression model. Then, forecast moving holiday ELD execution is performed using a constructed model to calculate the forecasting error using data in the evaluation part. Lastly, this study uses the constructed regression model to forecast moving holiday ELD for the year 2018. The error cannot be calculated in this stage because the actual data was not published yet on the TNB website.

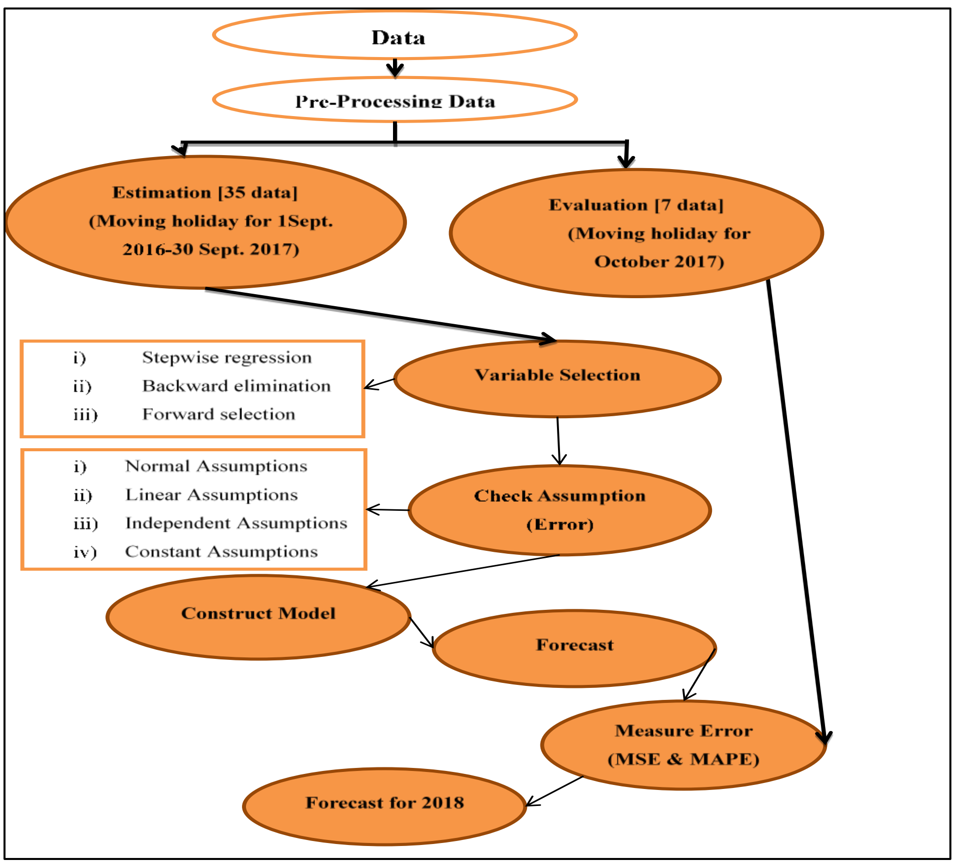

Fig 1. Framework of study

Data Analysis and Result

Multiple Linear Model

Table 2 gives the coefficient for all 12 independent variables and significant p-value. For the individual part, the p-value in the coefficient table looks significant for individuals or each variable. A significant independent variable will have a p-value of less than 0.05. From the result, the independent variables X_1, X_4, X_7, Z_1 and Z_5, were found to be significant. Based on Table 2, the estimated linear regression equation is given by

$Y = 7923.294 + 0.53X_1 - 0.225X_2 + 0.196X_3 - 0.54X_4 + 0.198X_5 - 0.117X_6 + 0.432X_7 - 172.736Z_1 - 1532.789Z_2 + 681.004Z_3 + 49.832Z_4 - 956.687Z_5$.

This model has the multiple correlation coefficient, R = 0.947, indicating a strong correlation between ELD and the one predicted by the regression model. The R Square value of 0.897 implies that all independent variables explain 89.7% variation in the dependent variable. The model formed from these variables is significant, where the p-value for this model is less than 0.05.

Table 2. Full model coefficients

Model	Unstandardized Coefficients		Standardized Coefficients	Test value	p-value
	B	Std. Error	Beta		
Constant	7923.294	3467.533		2.285	0.032
X_1	0.530	0.145	0.568	3.653	0.001
X_2	-0.225	0.174	-0.241	-1.296	0.208
X_3	0.196	0.194	0.196	1.007	0.325
X_4	-0.540	0.205	-0.421	-2.634	0.015
X_5	0.198	0.187	0.131	1.058	0.302
X_6	-0.117	0.206	-0.079	-0.568	0.576
X_7	0.432	0.140	0.301	3.082	0.005
Z_1	-1172.736	341.641	-0.383	-3.433	0.002
Z_2	-1532.789	781.831	-0.180	-1.961	0.063
Z_3	681.004	535.758	0.180	1.271	0.217
Z_4	49.832	501.722	0.016	0.099	0.922
Z_5	-956.687	277.392	-0.329	-3.449	0.002

Variable Selection Process

Stepwise selection

For stepwise selection, the model starts with an empty model. Then, the variable will be added one by one based on the smallest p-value from Table 2. When the variable is added to the model, they will be removed if the model is not significant after the variable's addition. It will stop until the model is significant, and if no variable satisfies, the entry criteria will be added to the model.

Table 3. Coefficients from Stepwise selection

Model	Unstandardized Coefficients		Standardized Coefficients	Test value	p-value
	B	Std. Error	Beta		
Constant	8737.833	2704.788		3.231	0.003
X_1	0.315	0.084	0.338	3.748	0.001
Z_1	-1144.094	261.671	-0.373	-4.372	0.000
X_7	0.440	0.112	0.307	3.920	0.001
Z_5	-869.734	226.072	-0.300	-3.847	0.001
X_4	-0.336	0.099	-0.262	-3.397	0.002
Z_2	-1551.929	663.492	-0.182	-2.339	0.027

Table 3 gives the final results based on the Stepwise variable selection process. From the results, six independent variables, X_1, X_4, X_7, Z_1, Z_2, and Z_5, are found to be significant. The estimated linear regression equation is given by

$$\hat{Y} = 8737.833 + 0.315X_1 -0.336X_4 + 0.44X_7 - 1144.094Z_1 - 1551.929Z_2 - 1551.929Z_5.$$

This model has the multiple correlation coefficient R = 0.933, indicating a strong correlation between electricity demand and those predicted by the X_1, X_4, X_7, Z_1, Z_2 and Z_5. Moreover, the R Square value is 0.87. This implies that all independent variables explain 87% of the variation in the dependent variables, which is X_1, X_4, X_7, Z_1, Z_2 and Z_5. The model formed from these variables is significant, where the p-value for this model is less than 0.05.

IV. BACKWARD ELIMINATION

For the Backward elimination process, the model starts with the full model. The variable that has the biggest p-value, as listed in Table 2, will be eliminated one by one. The process will stop until the model is significant, and those variables not satisfying the removal criteria will be removed from the model.

Table 4. Coefficient from the Backward elimination

Model	Unstandardized Coefficients		Standardized Coefficients	Test value	p-value
	B	Std. Error	Beta		
Constant	8992.434	2591.609		3.470	0.002
X_1	0.390	0.090	0.418	4.350	0.000
X_4	-0.376	0.097	-0.293	-3.877	0.001
X_7	0.394	0.110	0.274	3.570	0.001
Z_1	-1288.586	261.766	-0.421	-4.923	0.000
Z_2	-1490.996	635.687	-0.175	-2.345	0.027
Z_3	664.637	351.203	0.176	1.892	0.069
Z_5	-997.578	226.623	-0.344	-4.402	0.000

From the results in Table 4, only seven independent variables X_1, X_4, X_7, Z_1, Z_2, Z_3 and Z_5 are significant. The estimated linear regression equation based on Table 4 is given by

$\widehat{Y} = 8992.434 + 0.39X_1 - 0.376X_4 + 0.394X_7 - 1288.586Z_1 - 1490.996Z_2 + 664.63Z_3 - 997.578Z_5$.

This model has the multiple correlation coefficient R equal to 0.941, and the R Square is 0.885. This implies that this model explains about 88.5% of the dependent variables' variation, which is X_1, X_4, X_7, Z_1, Z_2, Z_3 and Z_5. The model formed from this variable is significant, where the p-value for this model is less than 0.05.

Forward Selection

For the Forward selection, the model starts with an empty model. Then, the variables were added one by one based on the smallest p-value in Table 2. It will stop until the model is significant, and no variable satisfying the entry criteria will be added to the model. Different from the Stepwise selection process, the Forward selection process does not execute the variable criteria after adding the variables. Table 5 shows the final results from the Forward selection process. The estimated linear regression equation is given by

$\widehat{Y} = 8737.833 + 0.315X_1 - 0.336X_4 + 0.44X_7 - 1144.094Z_1 -1551.929Z_2 -1551.929Z_5$,

as shown in Table 5. Only six independent variables, X_1, X_4, X_7, Z_1, Z_2 and Z_5, are significant. The multiple correlation coefficient for this model, R = 0.933, indicates a strong correlation between electricity demand and those predicted by the X_1, X_4, X_7, Z_1, Z_2 and Z_5. The R Square is 0.87, which means that all independent variables explain 87% of the dependent variables' variation, which is X_1, X_4, X_7, Z_1, Z_2 and Z_5.

Table 5. Coefficient for the Forward selection

Model	Unstandardized Coefficients		Standardized Coefficients	Test value	p-value
	B	Std. Error	Beta		
Constant	8737.833	2704.788		3.231	0.003
X_1	0.315	0.084	0.338	3.748	0.001
Z_1	-1144.094	261.671	-0.373	-4.372	0.000
X_7	0.440	0.112	0.307	3.920	0.001
Z_5	-869.734	226.072	-0.300	-3.847	0.001
X_4	-0.336	0.099	-0.262	-3.397	0.002
Z_2	-1551.929	663.492	-0.182	-2.339	0.027

Table 6 gives a summary of the results based on different variable selection techniques.

Table 6. Summary of final models based on variable selection process

Model	Significant variables	R	R square
Stepwise selection	X_1, X_4, X_7, Z_1, Z_2 and Z_5	0.933	0.87
Backward selection	X_1, X_4, X_7, Z_1, Z_2, Z_3 and Z_5	0.941	0.885
Forward selection	X_1, X_4, X_7, Z_1, Z_2, and Z_5	0.933	0.87

Table 6 shows the results to be similar to the results between Stepwise and Forward selection. This is because no variable met the removing criteria in the Stepwise selection process. In a nutshell, all variable selection processes choose the significant variables as X_1, X_4, X_7, Z_1, Z_2, and Z_5, but the Backward selection added another variable, which is Z_3. Therefore, we decided to select the best significant models considering variables X_1, X_4, X_7, Z_1, Z_2, and Z_5 only in the model. However, when involving dummy variables, it has to choose all of the dummy sets due to Z_2, Z_3 and Z_4 dummy variables representing a coincident variable. Thus, the significant variables for this model are X_1, X_4, X_7, Z_1, Z_2, Z_3, Z_4, and Z_5. We then proceed with the multiple regression analysis using X_1, X_4, X_7, Z_1, Z_2, Z_3, Z_4, and Z_5. The result of the model is given by

$$\widehat{Y} = 9022.898 + 0.393X_1 - 0.379X_4 + 0.396X_7 - 1318.59Z_1 - 1538.528Z_2 + 627.236Z_3 - 73.962Z_4 - 994.54Z_5,$$

with R equal to 0.941 and R square equal to 0.885. For the next section, the test of all assumption was analyzed based on this section's decision model.

Assumptions
Normal Assumptions

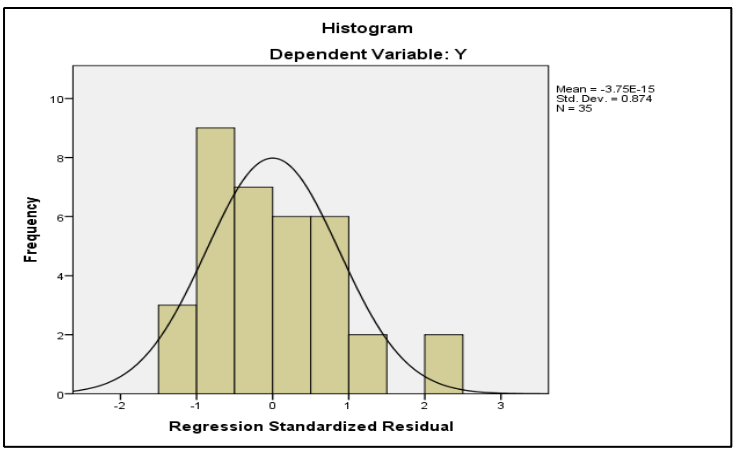

Fig 2. Histogram

Table 7. Test of normality result

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	p-value	Statistic	df	p-value
Unstandardized Residual	0.102	35	0.200	0.953	35	0.138
Standardized Residual	0.102	35	0.200	0.953	35	0.138

The histogram in Figure 2 shows a bell shape. So, it can be concluded from the histogram that they fulfilled the assumptions of normal distributions. Based on Table 7 of residual normality test, the p-value for unstandardized residuals and standardized residuals of Kolmogorov-Smirnov and Shapiro-Wilk would assume that the residuals were normally distributed because the p-value of 0.2 and 0.138 are greater than 0.5.

Linear Assumptions

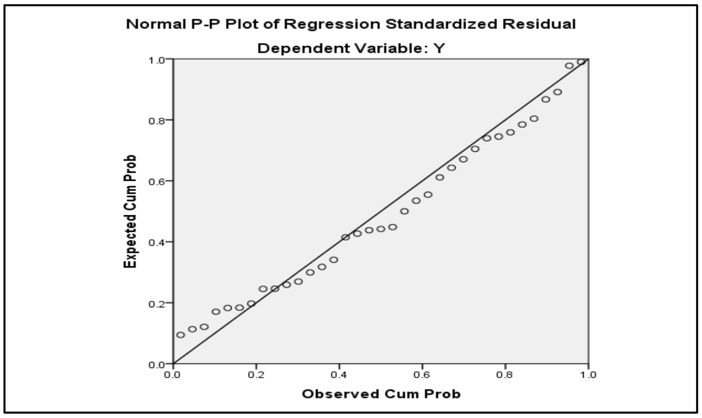

Fig 3. Normal P-P Plot of Regression Standardized Residuals

Figure 3 shows the normal probability plot for standardized regression residuals because the points more or less follow the straight line. There occur some deviation towards the center, but generally, the points seem to follow the line. Thus, it would assume linear distribution.

Independent Assumptions

Table 8. Durbin-Watson test results

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.941	0.885	0.850	558.892	2.093

Table 8 shows that the result of Durbin-Watson is 2.093, and the value is near 2, so we can conclude from the results that it fulfills the independent assumptions.

Constant Assumptions

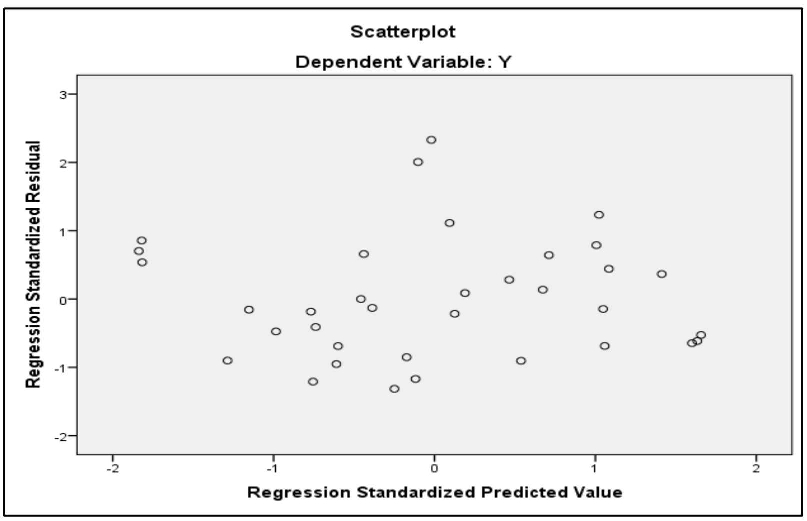

Fig 4. Scatterplot

Based on Figure 4, there is no point outside of negative 3 to 3, either at the x-axis or the y-axis. All scattered plots are distributed, and no pattern occurs. So, these scatter plots show that all residual fulfills the assumptions of constant variance.

After all the multiple regression analysis assumptions were proved met, the next section is to forecast moving holiday ELD using constructed model from the previous section.

Forecast for evaluation

Table 9 shows the actual and forecast value for a week of moving holiday in 2017, the Deepavali event (18th October 2017). From the calculation, the MSE is 1384742, MAPE is 6.62%, and APE on Deepavali moving holiday is 3.71%.

Table 9. Actual and forecast value for the week of Deepavali celebration

Date	X_1	X_4	X_7	Z_1	Z_2	Z_3	Z_4	Z_5	Actual	Forecast	APE
15/10/2017	15055	17126	14767	0	0	0	1	0	14983	14223.7	5.07
16/10/2017	14983	17082	16995	0	0	0	1	0	17034	15095.8	11.38
17/10/2017	17034	15575	17010	0	0	0	1	0	17255	16478.6	4.50
18/10/2017	17255	15055	17126	1	0	0	1	0	14936	15490.1	3.71
19/10/2017	14936	14983	17082	0	0	0	1	0	17135	15907.4	7.16
20/10/2017	17135	17034	15575	0	0	0	1	0	16809	15396.1	8.41
10/21/2017	16809	17255	15055	0	0	0	1	0	15951	14977.9	6.10

Forecast for next year moving holiday

In 2018, for big events such as Hari Raya AidilFitri and Chinese New Year, both estimates occurred on Friday, which is Hari Raya AidilFitri on 15th June 2018 and Chinese New Year on 16th February 2018. While conducting this research, there was no actual data for the year 2018. Therefore, we adopted ELD historical data similar to the big event that occurred on Friday to mimic the real data of 2018. This section adopts ELD data from 3rd to 9th December 2002, where Hari Raya AidilFitri occurs on 6th December 2002 (Friday). The result is shown in Table 10.

Table 10. Result of ELD forecast value on Chinese New Year and Hari Raya Aidilfitri 2018

Day	X_1	X_4	X_7	Z_1	Z_2	Z_3	Z_4	Z_5	Forecast
Tue	10506	10098	10231	0	0	0	1	0	13303.32
Wed	10399	9752	10408	0	0	0	1	0	13462.65
Thu	9981	8334	10507	0	0	0	1	0	13875.36
Fri	8037	10506	10098	1	0	0	0	0	10881.48
Sat	6396	10399	9752	0	0	0	1	0	11384.94
Sun	6493	9981	8334	0	0	0	1	0	11019.30
Mon	6694	8037	10506	0	0	0	1	0	12696.51

Another moving holiday in 2018 is Hari Raya AidilAdha and Deepavali, where both estimates will occur on Wednesday, Hari Raya AidilAdha on 22nd August and Deepavali on 7th November 2018. Similar to the previous forecast activities, while performing this research, there is no actual data in the year 2018. Therefore, we adopt ELD historical data that have similar big events on Wednesday as a mimic of the real data of 2018. This section adopts ELD data from 15th to 21st October 2017, where Deepavali occurred on 18th October 2017 (Wednesday). The result is shown in Table 11.

Table 11. Result of ELD forecast value on Deepavali and Hari Raya AidilAdha 2018

Day	X_1	X_4	X_7	Z_1	Z_2	Z_3	Z_4	Z_5	Forecast
Sun	15054.6	17126	14767	0	0	0	1	0	14223.70
Mon	14983	17081.6	16995	0	0	0	1	0	15095.82
Tue	17034	15575	17010	0	0	0	1	0	16478.59
Wed	17255	15054.6	17126	1	0	0	1	0	15490.12
Thu	14936	14983	17081.6	0	0	0	1	0	15907.37
Fri	17135	17034	15575	0	0	0	1	0	15396.11
Sat	16809	17255	15054.6	0	0	0	1	0	14977.92

V. CONCLUSION

The ELD forecasting plays an important role in capacity planning, scheduling, and operation of power systems. This study employs a multiple linear regression method in predicting the ELD. This method is valuable in economic and business research and helps establish functional relationships between two or more variables. In addition, it predicts the value of the dependent variable from the value of the independent variable, and it also tells the nature of the relationship. This regression method can also determine the model, the dependent variable and the potential independent variable. Using the regression method, this study achieved the objectives, which are to identify the significant variable that affects the ELD in the model and construct a multiple linear regression model for moving holiday ELD. Lastly, to forecast the ELD three days before the holiday, on the holiday and three days after the holiday for moving big holiday events, which are Hari Raya AidilFitri and Chinese New Year and not big events such as Hari Raya AidilAdha and Deepavali. Interestingly, the model constructed in this study can forecast ELD on the day of the moving holiday with the smallest forecast error of 3.7% compared to other days in the week. In general, this model may also be considered a good model with MSE equal to 1384742 and MAPE equal to 6.62% in the week of moving holiday. Future research suggests improving the model by considering more factors, such as temperature and school break factors. Other than that, combining and hybrid with other forecasting methods is also in progress, such as using fuzzy time series forecasting approach and combining the statistical forecasting model.

Reference

[1]. C. Viauroux, “Pricing urban congestion: A structural random utility model with traffic anticipation,” European Economic Review, vol. 55, no. 7, pp. 877–902, Oct. 2011.

[2]. I.Thomson and A. Bull, “Urban traffic congestion: Its economic and social causes and consequences,” CEPAL Review, vol. 2002, no. 76, pp. 105–116, Oct. 2006.

[3]. S. Ye, “Research on Urban Road Traffic Congestion Charging Based on Sustainable Development,” Physics Procedia, vol. 24, pp. 1567–1572, 2012.

[4]. F. Rempe, G. Huber, and K. Bogenberger, “Spatio-Temporal Congestion Patterns in Urban Traffic Networks,” Transportation Research Procedia, vol. 15, pp. 513–524, 2016.

[5]. T. Tettamanti and I. Varga, “Traffic control designing using model predictive control in a high congestion traffic area,” Periodica Polytechnica Transportation Engineering, vol. 37, no. 1–2, p. 3, 2009.

[6]. S. Petar and M. Ivaković-Babić, “Unification of Logistic Demands of Small-scale Enterprises as Solution of Urban Traffic Congestion Problem,” PROMET - Traffic&Transportation, vol. 23, no. 4, pp. 297–301, Jan. 2012.

[7]. S. V. Kumar and R. Sivanandan, “Traffic Congestion Quantification for Urban Heterogeneous Traffic Using Public Transit Buses as Probes,” Periodica Polytechnica Transportation Engineering, vol. 47, no. 4, pp. 257–267, Jan. 2018.

[8]. A.Choudhary and S. Gokhale, “On-road measurements and modelling of vehicular emissions during traffic interruption and congestion events in an urban traffic corridor,” Atmospheric Pollution Research, vol. 10, no. 2, pp. 480–492, Mar. 2019.

[9]. J. Qin, G. Mei, and L. Xiao, “Building the Traffic Flow Network with Taxi GPS Trajectories and Its Application to Identify Urban Congestion Areas for Traffic Planning,” Sustainability, vol. 13, no. 1, p. 266, Dec. 2020.

[10]. S. Zhang, S. Li, X. Li, and Y. Yao, “Representation of Traffic Congestion Data for Urban Road Traffic Networks Based on Pooling Operations,” Algorithms, vol. 13, no. 4, p. 84, Apr. 2020.

[11]. D. R. ALEKO and S. Djahel, “An Efficient Adaptive Traffic Light Control System for Urban Road Traffic Congestion Reduction in Smart Cities,” Information, vol. 11, no. 2, p. 119, Feb. 2020.

[12]. H. Karimi, B. Ghadirifaraz, S. N. Shetab Boushehri, S.-M. Hosseininasab, and N. Rafiei, “Reducing traffic congestion and increasing sustainability in special urban areas through one-way traffic reconfiguration,” Transportation, Jan. 2021.

[13]. X. Shi and Y. Zhao, “Traffic flow prediction model of urban traffic congestion period based on internet of vehicles technology,” International Journal of Information and Communication Technology, vol. 1, no. 1, p. 1, 2021.

[14]. L. Qi, M. Zhou, and W. Luan, “A Two-level Traffic Light Control Strategy for Preventing Incident-Based Urban Traffic Congestion,” IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 13–24, Jan. 2018.

[15]. A. Ali, M. A. Qureshi, M. Shiraz, and A. Shamim, “Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control,” Sustainable Computing: Informatics and Systems, vol. 32, p. 100608, Dec. 2021.

[16]. Y. Zhang, Y. Zhang, and R. Su, “Pedestrian-Safety-Aware Traffic Light Control Strategy for Urban Traffic Congestion Alleviation,” IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 1, pp. 178–193, Jan. 2021.

[17]. N. Wu, D. Li, and Y. Xi, “Distributed Weighted Balanced Control of Traffic Signals for Urban Traffic Congestion,” IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 10, pp. 3710–3720, Oct. 2019.

[18]. A. Choudhary and S. Gokhale, “On-road measurements and modelling of vehicular emissions during traffic interruption and congestion events in an urban traffic corridor,” Atmospheric Pollution Research, vol. 10, no. 2, pp. 480–492, Mar. 2019.

[19]. F. Grillo and J. Laperrouze, “Measuring the Cost of Congestion on Urban Area and the Flexible Congestion Rights,” Journal of Management and Sustainability, vol. 3, no. 2, Feb. 2013.

[20]. T. Osman, T. Thomas, A. Mondschein, and B. D. Taylor, “Does traffic congestion influence the location of new business establishments? An analysis of the San Francisco Bay Area,” Urban Studies, vol. 56, no. 5, pp. 1026–1041, Oct. 2018.