The relationship of economic variables and final energy consumption: multiple linear regression evidence

*L Abdullah**, and *W H Leong*

School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia

Abstract. Energy consumption in developing countries is sharply increasing due to higher economic growth of industrialization along with population growth and urbanization. This paper provides a multiple linear regression evidence to illustrate the association between final energy consumption and three economic variables. Multiple linear regression analysis was used to obtain a predictive equation and to check for linearity assumption. Three input variables, viz. growth domestic product, population, and tourism are the predictors for final energy consumption. Time series yearly data of final energy consumption and the three input variables for the year 2001 until 2012 was retrieved from various databases. It is found that there was a significant variation in the final energy consumption explained by the three variables. The multiple linear regression equation indicates that the ‘population’ is the most influential variable in predicting final energy consumption.

1 Introduction

It is undeniable fact that energy is one of the most basic needs in people’s lives and absolutely necessary to uplift the quality of life of the world’s population. The importance of energy has been accepted by all as energy uses to promote and leverages the world economically through many economic activities. There are several energy resources on earth likes fossil fuels (coal, natural gas and oil), hydroelectric power, biofuels, nuclear power, solar power, wind power, tidal power, wave power, geothermal power and others. These energy sources will ultimately consumed by end users. The total energy consumed by end users such as households, industry and agriculture is defined as final energy consumption. According to Energy Commission Malaysia [1], in the year 2013 alone, total energy consumption of Malaysia was 34586 ktoe. It was 25% higher than the consumption in the year 2012. Developing countries such as Malaysia is not exceptional in the need for energy. It is used to overcome the increasing demand of energy consumption, which is supposed to be moved in tandem with the growth of economic activities. Industrialization, population growth and urbanization are among the important variables that lead to expanding demand of energy [2]. Globally, energy demand seems to increase per annum due to anticipated higher Gross Domestic Product (GDP) growth. Apart from that, the

* Corresponding author: lazim_m@umt.edu.my

transportation and industrial will persist to be the major energy consumers by 41.1% and 38.8% respectively of the total energy demand by 2010 in Malaysia [2]. In other words, energy consumption and economic activities are mutually related and plays a very important role in the economic growth and social development of a country.

It is believed that there are many variables contribute to the final energy consumption. Zhou et al., [3] and Kuo et al.,[4] suggested the variables that can be linked to final energy consumption are manufacturing, residential household, agriculture, industrial, commercial, construction, services, and transportation, just to name a few. Furthermore, Zhou et al., [3] mentioned the macroscopic variables such as GDP, population and level of urbanization may influence final energy consumption. In searching for other variables, Kou et al.,[4] also included numbers of tourist arrivals as another variable. Some related researches investigate a single variable that may influence final energy consumption. For example, Zhong [5] suggests single variable, namely GDP affects the final energy consumption in China. It is believed that many other variables can also be related to final energy consumption. However, the contribution of each variable towards final energy consumption is hardly be measured. The conclusive association between economic variables and final energy consumption has always become a crucial issue among policy makers. The present study investigates the relationship between three economic variables and final energy consumption using multiple linear regression analysis. A multiple linear regression equation is a linear association between one dependent variable and at least two independent variables.

2 Related research

Many studies have been conducted to investigate the relationship between variables and energy consumption using wide ranges of approaches from typical statistical methods to intelligent based methods. Kucukali and Baris [6], for example, introduced the intelligent method fuzzy logic to forecast Turkey's short-term gross annual electricity demand. GDP based on purchasing power parity was the only parameter used in the model. The results show the electricity demand is strongly related with the GDP. In another intelligent based method, Zhong [5] constructed a grey dynamic model to predict the clean energy consumption in China. In the study, the data from year 1995 to the year 2011 was used to make a prediction. The independent variables were GDP and effluent charge, while dependent variable was clean energy consumption. The result showed that GDP and effluent charge affect the clean energy consumption in China, but GDP less affected to the clean energy consumption when compared to effluent charge. Bazmi et al.,[7] developed a prediction for electricity demand by using Adaptive Neuro-Fuzzy Network in a case study in Johor, Malaysia. In this case study, they designed an adaptive neuro fuzzy inference system (ANFIS) network to map six independent variables as input variable to electricity demand as output variable. The input variables of the study were economic, demographic and meteorological parameters. Specifically, GDP, employment, industry efficiency, population, minimum average annual temperature and maximum average annual temperature were the input variables.

Besides using the intelligent based methods, there are also a comparative study between intelligent based methods and statistical analysis. Maliki et al.,[8] conducted a comparative study to see the performance of Artificial Neural Network (ANN) and Regression method to forecast the electrical power in Nigeria. In the study, Instantaneous Annual Peak Load Demand and Annual Average Load Demand as input variables (independent variable) whereby the estimated electrical Power Generated as output variable. They compared the performance of ANN and Regression method using Root Mean Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MAE). The result shows that ANN is

a better way in forecasting. In a time series analysis, Kuo et al.,[4] analysed the relationship between tourism development, economic growth, CO2 emission and energy consumption using Autoregressive Integrated Moving Average Model (ARIMA). The result indicates that an increase of tourist arrival would raise the energy consumption and emission of CO2, while the increase of visitors results in the rise of GDP. Braun et al., [9] estimated the energy consumption by using Multiple Regression Analysis (MRA). A case study of a supermarket in northern England, United Kingdom was considered in the study. They used two separable input variables which were dry-bulb temperature and relative humidity while energy consumption as output variable. Two data sets from division of input variables were used to generate two equations (electricity and gas) by MRA. The energy consumption from 1961 to 1990 was estimated using the two equations of MRA. Differently from the above related research, this study aims to unravel the relationship between three economic based variables and final energy consumption using multiple linear regressions.

3 Research framework

In order to achieve the set out objective, three input variables and one output variables are identified. The GDP, population and tourism that represent economic variables are chosen as input variables while final energy consumption as output variable. The choices of these variables are in accordance with several previous studies [4],[5]. This study employs secondary data collected from World Bank [10],[11], Tourism Malaysia [12], Department of Statistics [13] and Energy Commission [1] websites. Twelve-yearly data from 2001 until 2012 pertaining to the input and output variables are carefully collected and analyzed using multiple linear regressions.

The goal of multiple linear regressions is to allow a researcher estimates the relationship between a dependent variable and several independent variables. Dependent variable also named as predicted while independent variable named as predictor. The end result of multiple linear regression is the development of a regression equation (line of best fit) between the predicted and predictors. The general equation for regression equation can be written as

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1x_{i,1} + \hat{b}_2x_{i,2} + \hat{b}_3x_{i,3} + ... + \hat{b}_4x_{i,K} + \varepsilon_i \quad (1)$$

The constant $\hat{b}_n$ is called regression coefficients or regression weights. The symbol “^” represents predicted values and ε_i is a random error term with constant variable and zero mean, while $i = 1, 2, 3, ..., n$. The errors are assumed independent and follow standard normal distribution $N(0, \sigma^2)$. However, there are several assumptions related to linear regression which are the y values are statistically independent. Also, there is a group of y values are normally distributed for each value of x . All the mean of normal distributions of y values lie on the line of regression and the standard deviations of normal distributions are equal.

4 Analysis and results

It is recalled that the main objective of this analysis is to establish the association between selected economic variables and final energy consumptions. The analyses are performed according to the following steps.

Step 1: Define the input variables and output variable.

A Multiple linear regression equation is a linear association between one dependent variable (y) and at least two independent variables (x_n). Therefore, input variables and output variable are defined as,

- Output Variable: Final energy consumption (y)
- Input variables: GDP (x_1), Population (x_2), and Tourism (x_3).
- Step 2: Collect and key in the data of input variables and output variable.

The yearly data of final energy consumption, GDP, population and tourism of Malaysia from year 2001 until 2012 were collected and analyzed with the help of a statistical software.

Step 3: Compute and interpret the linear relationship between input variables and output variable.

It is assumed that there is an appropriate model in relation to the relationship between variables. The must know term in regression analysis is multiple correlation coefficient (R^2) which is a statistical measure to show how close the data with the fitted regression line [14]. It is also known as coefficient of multiple determinations. Using the collected data, multiple correlation coefficient (R^2) of the model is 0.932. It indicates that 93.2% of the variation in the final energy consumption is explained by GDP, population and tourism. Since R^2 is close to 1, therefore the regression equation appeared to be useful for estimation.

Step 4: Test the null hypothesis using F-test, to check whether there is a regression relation between GDP, population, tourism and final energy consumption.

The hypothesis of the multiple linear regression model is stated as follows.

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_1 : \text{at least one } \beta_i \neq 0$$

(2)

The result of F-test is shown in table 1.

Table 1. F-Test for Regression Analysis.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	201231528510008.800	3	67077176170002.940	36.692	0.000
Residual	14624740489991.172	8	1828092561248.897		
Total	215856269000000.000	11			

It is shown that the value of F is 36.692 and p -value is 0.000. Since p -value is less than 0.05 of significance level, α , then the null hypothesis is rejected. At the 5% significance level, there is enough evidence to statistically prove that at least one of the input variables is useful in predicting the output variable.

Step 6: Determine the regression coefficients, b_n and the multiple regression equation. The end result of multiple linear regression is the development of a regression equation (line of best fit) between the predicted and predictors. The analysis is summarized in table 2.

Table 2. Unstandardised Coefficients.

Input variable	Unstandardised coefficients		Standardised coefficients
	B	Std error	Beta
Constant	-26090248.469	17510374.889	
GDP	3.727×10^{-6}	0.000	0.259
Population	2.543	0.839	1.000
Tourism	-0.247	0.387	-0.290

Based on the information from table 2, the regression equation of the model is written as

$$\hat{Y}_i = -26090248.469 + 0.000003727x_1 + 2.5433x_2 - 0.247x_3 \quad (3)$$

The equation clearly indicates that the variable population is the most influential variable in predicting the final energy consumption in Malaysia. It can conclude that multiple linear regression model has performed well in predicting the final energy consumption in Malaysia. Hence, the economic variables viz GDP, population and tourism are contributing to the final energy consumption.

5 Conclusion

The relationship between energy consumption and its related multiple variables is a complicated issue in which a multivariate model must be considered. The purpose of the current study was to determine the relationship between final energy consumption and its multiple economic variables using multiple linear regressions. Three variables were identified as the input variables that could be used to predict the final energy consumption. The eight steps of multiple linear regression analysis were employed in this investigation. This study found that generally, all the three variables, viz., GDP, population and tourism were contributed to the final energy consumption. However, the major finding was the establishment of the most influential factor to the final energy consumption. The result suggests that the most influential variable in final energy consumption is population. This research has served as a basis for future studies where more input variables could be taken in account together with longer time series data. The research was limited in several ways. First, the research used a secondary data that believed to be accurate and trusted. The method used was a linear model in which a linearity assumption was tested. This linearity assumption may undermine the nature of secondary data which sometimes exhibits non-linear characteristics. This is just a piece of preliminary study where further experimental investigations are needed in future particularly in comparing the accuracy of the multiple linear regression model with other non-linear models.

References

1. Energy Commission. (2014). National energy balance. 2012<http://meih.st.gov.my/publications>. [8 December 2014].
2. United Nations Development Programme (UNDP). (2006). Achieving industrial energy efficiency in Malaysia. ISBN 983-40995-7-6.
3. Zhou, W., Zhu, B., Chen, D., Ma, Y. & Fei, W. (2011). Energy consumption patterns in the process of China's urbanization. *Population Environment* 33, 202–220.
4. Kuo, K. C., Liu, M. & Lai, S. L. (2012). Effect of tourism development on energy consumption, CO2 and economic growth in China. *Advanced Materials Research*, 10, 524-527.
5. Zhong, X. W. (2015). A Predictive analysis of clean energy consumption, economic growth and environmental regulation in China using an optimized grey dynamic model. *Computational Economics* 46, 1-17.
6. Kucukali S. & Baris, K. (2010), Turkey's short-term gross electricity demand forecast by fuzzy logic approach, *Energy Policy*, 38, 2438-2445.
7. Bazmi, A. A., Davoody, M. & Zahedi, G. (2012). Electricity demand estimation using an adaptive neuro-fuzzy network: a case study from the state of Johor, Malaysia, *International Journal of Chemical and Environmental Engineering* , 3(5), 284-292.

8. Maliki, O. S. , Agbo, A. O., Maliki, A. O., Ibeh, L. M. & Agwa, C. (2011). Comparison of Regression Model and Artificial Neural Network Model for the prediction of Electrical Power generated in Nigeria. *Advances in Applied Science Research*, 2(5), 329-339.
9. Braun, M. R., Altan, H. & Beck, S. B. M. (2014). Using regression analysis to predict the future energy consumption of a supermarket in the UK. *Applied Energy* , 130, 305-313.
10. The World Bank Group. (2015). GDP (Current U. S. dollars). <http://data.worldbank.org/indicator/NY.GDP.MKTP.CD?page=2>.
11. The World Bank Group. (2015). Population, total. <http://data.worldbank.org/indicator/SP.POP.TOTL>.
12. Tourism Malaysia. (2008). Tourist arrivals & receipts to Malaysia. http://corporate.tourism.gov.my/research.asp?page=facts_figures.
13. Department of Statistics, Malaysia. (2013). Annual Gross Domestic Product 2005–2012. <http://www.epu.gov.my/en/national-accounts>. [15 May 2013].
14. Jim, F. (2013). Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit? <http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>. [30 May, 2013].