KKU ENGINEERING JOURNAL 2016;43(S1):92-95 Research Article

KKU Engineering Journal

https://www.tci-thaijo.org/index.php/kkuenj/index

Using generalized linear models and time series models to forecast gasohol consumption in Thailand

Sujitta Suraphee[1], Weerapat Sessomboon[2] and Rojanee Homchalee*[1]

[1)]Research Unit on Applied Statistics, Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham 44150 Thailand.
[2)]Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002 Thailand.

Received April 2016
Accepted June 2016

Abstract

This paper proposed two methods in statistical modeling to forecast gasohol consumption as alternative energy in Thailand. The generalized linear models (GLMs) and time series models were applied. The first method was used to investigate the factors that impact response variables, consumption of gasohol 91, gasohol 95, gasohol E20, and gasohol E85. Also, the first method was to take among of factors to forecast all of response variables. The second method was used to forecast gasohol consumption in the further. The important results were ethanol price impacting to all of response variables: higher ethanol price making smaller consumption of gasohol 91 and gasohol 95. On the other hand, higher ethanol price making higher consumption of gasohol E20 and gasohol E85. However, it has been found in this study that forecast precision in terms of MAPE of GLMs models were less than time series models.

Keywords: Alternative energy, Gasohol, Generalized linear models, Time series model

1. Introduction

Thai government has continuously approved measures in order to support energy security of the country. In 2008, The Alternative Energy Development Plan (AEDP 2008-2022) was set. Ethanol is one of alternative energy sources in this plan and it has been used for blending the gasohol as biofuel instead of gasoline[1]. In later year 2012, The Renewable and Alternative Energy Development Plan for 25% in 10 Years (AEDP 2012-2021)" was set. In this plan, the consumption targets of ethanol should be 9 million liters per day within 2021 [2]. Gasohol consumption is increasing continuously, especially, gasohol 95 and gasohol 91. Total gasohol consumption in 2014 was 9,130.87 million liters. Currently, the new plan: "AEDP 2015-2036" has been set, with the consumption target of ethanol should be 11.3 million liters per day within 2036 [2]. In order to activate the plan, consumption of gasohol E20 and gasohol E85 should be increased. However, there are many factors effecting gasohol consumption such as petroleum price and other gasoline prices. Therefore, this study aims to find factors effecting gasohol consumption and formulate appropriate models by using generalized linear models and time series model to forecast gasohol consumption in the further, which is useful for government planning.

2. Materials and methods

2.1 Generalized linear models (GLMs)

Generalized linear models extend the linear modeling framework in case of response variables are not normally distributed. There are some situations where general linear models are not appropriate such as–the value of response variables is a discrete or it is continues variable which is not normality. The purpose of both linear models (LMs) and generalized linear models (GLMs) is to express the relationship between an observed response variable, *Y*, and a number of covariates (also called predictor or explanatory variables), *X*. Both models view the observations, Y_i, as being realizations of the random variable *Y*. The variable *Y* in linear model is considered as the sum of its mean, μ, and a random variable, ε [3]:

$$Y = \mu + \varepsilon ; \mu = E(Y) : \text{the expected value of } Y. \quad (1)$$

It is assumed that the expected value of *Y* can be written as a linear combination of the covariates, *X*, that is $\mu = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k$ and the error term, ε, is normally distributed with mean zero and constant variance making *Y* is also normal.

Usually, estimation of $\alpha, \beta_1, \ldots, \beta_k$ which is said parameters in the model can be easily used classical least square method. However, the required assumptions are not

*Corresponding author. Tel.: +6681 974 0250
Email address: rojaneeh@hotmail.com
doi: 10.14456/kkuenj.2016.79

easy to guarantee in applications. It is difficult to assert normality and constant variance for response variables. The values for the response variable may be restricted to be positive such as consumption of gasohol. The assumption of normality violates this restriction.

If the response variable is strictly non-negative then intuitively the variance of *Y* tends to zero as the mean of *Y* tends to zero. That is, the variance is a function of the mean. This situation should be applied by GLMs to construct suitable statistical modeling and maximum likelihood estimator is applied to estimate the parameters in the model. Distribution of the response variable in GLMs will be a member of the exponential family distributions [4]. Also, the variance is permitted to vary with the mean of the distribution. Finally, the effect of the covariates on the response variable is assumed to be additive on a transformed scale.

2.1.1 Components of a generalized linear model

- *Random component:* Each component of *Y* is independent and is from one of the exponential family of distributions.

- *Systematic component:* The *p* covariates are combined to give the linear predictor

- *Link function:* The relationship between the random and systematic components is specified via a canonical link function, *g*, that is differentiable and monotonic. The model formula states that

$$g(\mu) = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad (2)$$

2.1.2 The GLMs formula with Gamma distribution

Since the response variables in this study are consumption of gasohol 91, gasohol 95, gasohol E20, and gasohol E85, which are nonnegative values and not normality, therefore, we apply GLMs with Gamma distribution to construct the appropriated model. The canonical link hears is $g(\mu) = \mu^{-1}$;

$$\mu^{-1} = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k \quad (3)$$

2.2 Time series models

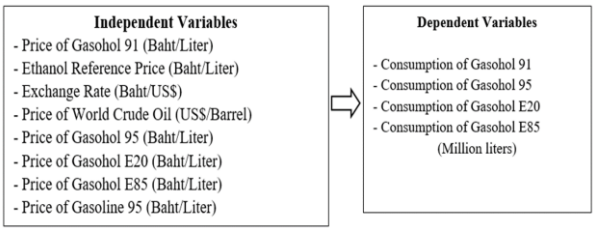
A time series data is a sequence of data that are arranged according to the time of their outcome. The objective of time series analysis is to develop mathematical models that provide descriptions for data, like that found in the previous part. Time series models for forecasting can be created from several techniques such as moving average, exponential smoothing, Box-Jenkins technique, and etc. An autoregressive integrated moving average (ARIMA) model is created from Box-Jenkins technique. The ARIMA is quite effective in analyzing the model than others, therefore, it is used to formulate model for forecasting [5].

Non-seasonal ARIMA models are normally indicated *ARIMA (p, d, q)* where parameters *p* is the order of the autoregressive model, *d* is the degree of differencing, and *q* is the order of the moving-average model. Seasonal ARIMA models are generally indicated *ARIMA (p, d, q)(P, D, Q)*$_m$, where *m* refers to the number of periods in each season, and *P, D, Q,* refer to the autoregressive, differencing, and moving average terms for the seasonal part of the model [6].

Time series analysis was used to formulate models for forecasting in many researches. For example, exponential smoothing and Box-Jenkins technique were used to formulate models for forecasting of ethanol production and consumption in Thailand. It was found that ARIMA or Box-Jenkins models were appropriate models, and could be used to determine appropriate quantity and proportion of feedstocks needed for ethanol production in the future [7].

From the literatures, two methods, GLMs and ARIMA are used to construct the forecasting models. After that the two types of forecasting models were compared by MAPE and the appropriate models to forecast the consumption of gasohol in Thailand were proposed.

2.3 Dependent and independent variables



3. Results

Consider from the mean absolute percent error (MAPE), the results found that the time series model is more appropriated than using the generalized linear model to forecast consumption of gasohol 91, gasohol 95, gasohol E20, and gasohol E85, as shown in Figure 1. However, they have different aims, the objective of using the GLMs is to take the affecting factors to forecast the response variables but the times series models use data in the past for forecasting data in the future.

3.1 The factors impact to gasohol consumption by using GLMs

We found that the distribution of consumption of gasohol is Gamma distribution. The GLMs formula with Gamma distribution was applied. The predictor variables which were significantly impact to the consumption of gasohol are as the following:

1. The consumption of gasohol 91 depends on ethanol price, price of gasoline octane 95 and price of world crude oil. If all of factors are increased, the consumption of gasohol 91 is decreased.

2. The consumption of gasohol 95 depends on ethanol price, price of gasoline octane 95, price of gasohol 95, price of gasohol 91, and price of world crude oil. Almost affecting factors are increased the consumption of gasohol 95 is decreased excepting price of gasohol 91.

3. The consumption of gasohol E20 depends on ethanol price and exchange rate. If both of affecting factors are increased, it will also make increasing of gasohol E20 consumption.

4. The consumption of gasohol E85 depends on ethanol price, exchange rate, and price of gasohol E85.If all of affecting factors are increased, it will also make increasing of gasohol E85 consumption.

3.2 The suitable time series models to forecast gasohol consumption

Gasohol 91: *ARIMA*(2,1,1)

$$(1+1.1294B+0.2778B^2)(1-B)\ln\hat{Z}_t = 0.0155+(1+0.9672B)\hat{a}_t \quad (4)$$

Gasohol 95: *ARIMA*(1,2,0)

$$(1+0.4216B)(1-B)^2\ln\hat{Z}_t = \hat{a}_t \quad (5)$$

Gasohol E20: *ARIMA*(4,2,1)

$$(1+1.4357B+1.3552B^2+0.9931B^3+0.4850B^4)(1-B)^2\ln\hat{Z}_t = (1+0.4679B)\hat{a}_t \quad (6)$$

Gasohol E85: *ARIMA*(2,2,0)

$$(1+0.7961B+0.3790B^2)(1-B)^2\ln\hat{Z}_t = \hat{a}_t \quad (7)$$

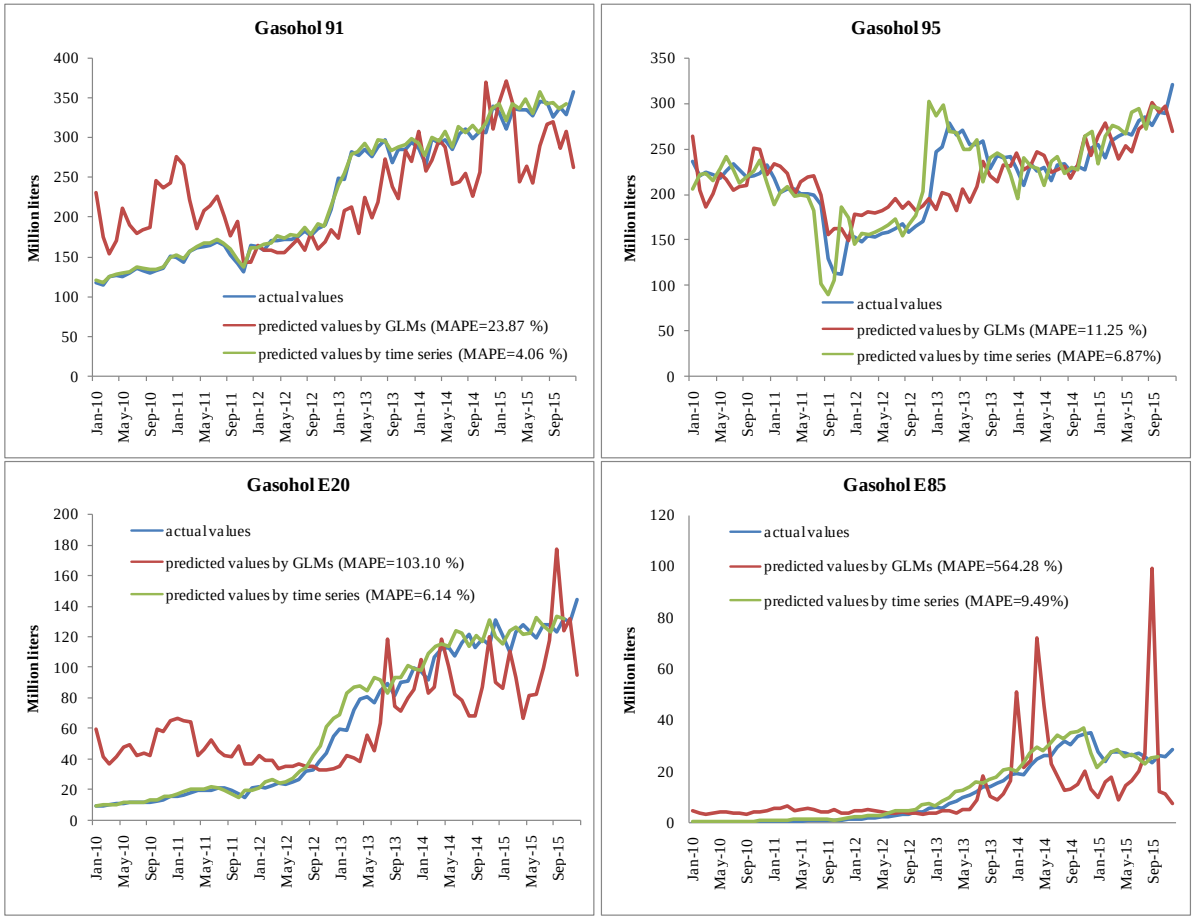


Figure 1 Actual and predicted value from GLMs and time series models

4. Discussion and conclusions

This paper proposes two methods in statistical modeling to forecast gasohol consumption in Thailand. The generalized linear models (GLMs) and time series models were applied. The GLMs is applied since the authors would like to find the factors that affect to consumption of gasohol in which it is not normally distributed. The important result is that ethanol price is the only one factor that affects to all of response variables. This result can be considered to increase material products of making the ethanol for Thai government.

For time series models, the ARIMA is quite effective in analyzing the model than other time series models because of the fact that there is a study on the dynamic relationship between variables and other components of the equation. Therefore, it is used to formulate model for forecasting gasohol consumption.

To forecast consumption of gasohol 91, gasohol 95, gasohol E20, and gasohol E85, the MAPE is considered. The results found that the time series models were more appropriated than using the generalized linear models since consumption of gasohol is a time series data which is an autocorrelation variable.

However, they have different aims, the objective of using the GLMs is to take the affecting factors for forecasting the response variables but the times series models use data in the part for forecasting data in the future.

5. Acknowledgements

The authors would like to thank Faculty of Science, Mahasarakham University for the financial support of this study.

6. References

[1] Department of Alternative Energy Development and Efficiency. Alternative Energy Development Plan in 15 Years (AEDP 2008-2022). Thailand: Department of Alternative Energy Development and Efficiency, Ministry of Energy; 2008.

[2] Department of Alternative Energy Development and Efficiency. The Renewable and Alternative Energy Development Plan for 25% in 10 Years (AEDP 2012-2021). Bangkok: Department of Alternative Energy Development and Efficiency, Ministry of Energy; 2012.

[3] Duncan A, Sholom F, Claudine M, Doris S, Ernesto S, Neeza T. A Practitioner's Guide to Generalized Linear Models. Virginia: Towers Watson; 2007.

[4] Nelder JA, Wedderburn RWM. Generalized Linear Models: Reviewed work(s). Journal of the Royal Statistical Society, Series A (General) 1972;135:370-384.

[5] Fildes R, Makridakis S. The Impact of Empirical Accuracy Studies on Time Series Analysis and Forecasting. International Statistical Review1995;63: 289-308.

[6] Bowerman BL, O'Connell RT. Forecasting and Time Series: An Applied Approach. Belmont, CA: Duxbury Press; 1993.

[7] Sriyotha S, Homchalee R, Sessomboon W. Forecasting of Production and Consumption of Ethanol in Thailand using Time Series Analysis. Applied Mechanics and Materials 2015;781:651-654.