Forecasting the fuel consumption based on the fuzzy linear regression models

Nadia Rasouli
Polytechnique Montréal

ISSN: 2456-043X

International Journal of Engineering Science and Generic Research (IJESAR)
Available Online at www.ijesar.in
Journal Index In ICI World of Journals - ICV 2016 – 68.35
Volume 4; Issue 3; May-June; 2018; Page No. 41-59

Forecasting the fuel consumption based on the fuzzy linear regression models

Nadia Rasouli

PhD. Candidate, Industrial Engineering, Tarbiat Modares University, Tehran, Iran.

Abstract

The growth and survival of most economic activities in developing countries depend on the energy supply issue. Therefore, government authorities try to favorably control energy supply and demand parameters by predicting energy consumption and proper planning to guide consumption as carefully as possible. Crude petroleum and petroleum products form a major part of energy carriers in Iran and thus the experts have had to take essential steps towards forecasting and planning in this regard. In this work, the demand for petroleum products including gasoline, gasoil, fuel oil, kerosene and LPG have been forecasted by linear regression, regression using the Cobb-Douglas function and fuzzy linear regression models considering the effective parameters. Data for the years 1996-2011 period have been used for the assessment of regression model and the 2012-2016 data have been used to select the appropriate model and the check its validity. Finally, the demand for petroleum products in 2017-2036 has been forecasted using the most appropriate model.

Keywords: Forecasting; Petroleum products; Fuzzy linear regression; Cobb-Douglas function; linear regression function.

1. Introduction

The twentieth century is referred to as the human significant progress era in all fields. Obviously, if all the activities in different sections of the country are not performed based on sound and rational planning, huge amounts of energy will be wasted. One of the most important viewpoints that tries to minimize the amount of wasted time and fuel is minimizing the traveled distance from the starting point to the desired destination. Many researchers have conducted interesting researches in this area. Two powerful instances in different sub-categories are (1) in exact algorithms and (2) in heuristic approximate algorithms. The first paper is very fast and exact algorithm that provides the mathematically global optimal solution which can be used in different work that need to be done swiftly in presence of convex and nonconvex polygonal obstacles but the second work comes up with a local optimal one. They are suitable for deterministic area but when there is uncertainty the planning requires accurate forecasting of energy consumption. In fact, the analysis of effective factors on energy consumption and a view of the prediction of consumption enable the mangers to take necessary measures for controlling energy supply and demand variables.

Energy sector is considered the infrastructure for development given its dual role in the energy supply and foreign exchange income and has had a fundamental part in the social/economical parts of the country. There is a significant relationship between the energy supply and foreign exchange income as indicated by the statistics. Addressing the consumption of petroleum products in different sections, which accounts for 46.1% of the entire energy carriers, is of great importance. Therefore, the analysis of the effective factors on the intensity of increased consumption and a view of the predicted energy consumption give managers the opportunity to take the required measures to control the energy supply and demand variables. Different technical and statistical methods have been proposed for forecasting energy consumption in the last few decades giving different results. However, no technique or combination of techniques has been sufficiently successful in forecasting energy consumption. Different models such as time series (3), econometrics (4), fuzzy logic (5), neural networks (6), genetic programming (7) and Particle Swarm Optimization (8, 9) have thus far been used for forecasting energy demand. Meta heuristic are used in different aspects including. Different studied are performed in energy planning such as (10, 9) and (11).

The objective of this work was forecasting the demand in a long term period and regression models are more highly capable of such predictions. Therefore, linear regression, regression using the Cobb-Douglas function and fuzzy linear regression models have been investigated for predicting the demand for petroleum products in this

Corresponding author: Nadia Rasouli

work and fuzzy linear regression model has been selected as the best model. Model inputs or the effective parameters on the demand for petroleum products were selected considering the studies carried out and the experts' opinion. The data for fuzzy regression model are certain variables and are fuzzy model parameters. This model has been used given that the inputs and output of the problem are certain. The results obtained are applicable for proper planning for the supply of petroleum products. Linear regression model has been proposed by Tanaka (12).

In the ordinary regression model, the objectives of regression analysis are: 1. finding the most appropriate regression model. 2. Determination of the best coefficients considering the inputs. In fact, the application of regression model is limited by the certain assumptions about the data. The model can only be used when the data have a statistical distribution and the relationship between x and y is a definite number. To overcome such constraints, fuzzy regression models have been proposed in order to use in the assessment of the variables with limited and imprecise available data and uncertain, qualitative and fuzzy relationships as well as improve the conventional regression models. One of the major differences between conventional and fuzzy linear regression models is that the difference between the observed and estimated values are associated with visual errors in the conventional regression while this difference is a result of inherent uncertainty in fuzzy linear regression. Fuzzy regression can be used for the assessment of both fuzzy and certain data while conventional regression is only applied in the assessment of certain data. The main objective of fuzzy linear regression model is finding a regression model for all the data observed according to the proportionality criterion considered.

The demand for petroleum products in 2017-2036 has been forecasted in this work. In order to forecast the demand for gasoline, the effective parameters including the number of cars, motorcycles, gross domestic product and population have been considered. The effective parameters on the demand for gasoil include gross generation by power plants within the past year; gross domestic product, population and the number of gasoil consuming vehicles. The effective parameters on the demand for fuel oil include gross generation by power plants within the past year, the consumption of energy carriers in industries within the past year; gross domestic product and population and the corresponding parameters for the demand for kerosene are population and gross domestic product. Finally, the effective parameters on the demand for LPG are the share of petroleum products in the domestic sector, population and gross domestic product. The data from 1996-2011have been used to evaluate forecasting models (conventional regression, fuzzy linear regression and regression using Cobb-Douglas function) for each of the petroleum products and the data from 2012-2016 have been used in the investigation the effectiveness and credibility of the models. Finally, the demand for petroleum products in 2017-2036 was forecasted using the most appropriate model.

In this work, a background of the models used to forecast energy demand will be given in the next section. Following the brief introduction of fuzzy regression model in section 3, the dependent and independent variables will be discussed in section 4. Conventional regression, fuzzy linear regression and regression using Cobb-Douglas function will be dealt with in section 5. The three forecasting models will be analyzed and compared in section 6 and the demand for petroleum products in 2017-2036 will be forecasted using the selected model in section 7. Finally, conclusions and suggestions for future research will be given in section 8.

2. Research Background

Literature survey shows that fuzzy regression models have been successfully applied in various fields of engineering and forecasting (13-16) as well as different applied areas such as marketing, energy and sale (13). These models are also used for forecasting energy demand. A variety of researches have been carried out for forecasting energy demand using these models. Al-Kandari et al. (2004) have developed a fuzzy regression model for forecasting electricity demand in winter and summer. Fuzzy estimation was used using linear optimization model in this research and forecasting electricity demand up to 14 hours is feasible(17). Taghizadeh et al. (2008) have modelled the forecasting of energy demand in Iranian transport section using multi-layer fuzzy linear regression. The model inputs were population, gross domestic product the number of vehicles and energy consumption by the transport section in the previous year and the output is energy consumption in the transport section. The data from 1993-2005 was used for modelling and checking the validity of the model used and the energy demand of the transport section in 2006-2020 has been forecasted(18). Shakouri and Nadimi (2009) forecasted the energy demand of the household sector using a novel fuzzy linear regression model. The model inputs were gross domestic product, the number of families and energy cost index and the output was offered in fuzzy form(19). Azadeh et al (2010) have proposed fuzzy regression model for estimation of crude oil consumption. Economic standard indices including population, crude oil import cost, gross domestic product and annual petroleum production in the past year were considered in this research. To investigate the efficiency and check the validity of the model, petroleum consumption data in Canada, the United States, Japan and Australia in 1990-2005 were used. They developed a fuzzy regression model to forecast electricity consumption in Iran (20).

Examples of energy forecasting problems using different models will follow. Al-Omishy (1989) developed a computer simulation model for forecasting gasoline and gasoil consumption by vehicles. The simulation model proposed provides the data formation of regression equations corresponding to fuel consumption considering the speed, type and duty cycle of the vehicle. The vehicles selected in this study were investigated in a wide range of traffic conditions. This model has also been designed for forecasting fuel consumption by vehicles on the road as well as the total fuel consumption(21).

Nizami and Al-gami (1995) have used a two layered feed forward neural network to study the relationship between electrical energy consumption in eastern states of Saudi Arabia and weather data, comprehensive radiation and population and have used the data from 7 years to train this network(22). Kalogirou (2000) has used artificial neural network technique for estimating thermal energy consumption in buildings and forecasting energy consumption in a passive solar building. A model has been proposed for forecasting energy consumption in passive solar building based on artificial neural network. They used back-propagation algorithm for training the network(23).

Kermanshahi and Iwamiya (2002) have proposed an artificial neural network for forecasting peak electrical load in Japan. Two neural networks including a three layer back propagation and a recurrent neural network have been used in this investigation. Ten factors have been considered as inputs of the network to forecast the peak electrical load and the forecasting was ultimately carried out up to 2020(24).

Nasr et al. (2003) proposed s neural network model for forecasting gasoline consumption in Lebanon. Four neural networks were proposed in this study. The first model was based on previous gasoline consumptions and the second model was a multi-variable model based on the time series of gasoline consumption and its price. The third model was multi-variable model based on time series of gasoline consumption and registration vehicle. Finally, the fourth or last model was a hybrid model based on gasoline price, time series of gasoline consumption and registration vehicle. The four models were studied based on different performance criteria and the fourth model was selected as the most efficient model(25).

Murat and Ceylan (2006) proposed a model based on neural network for forecasting energy demand and transport in Turkey. This model was based on supervised neural network approach and the corresponding social-economical and transport indices were used as inputs. The neural network model was a feed forward model with propagation training algorithm. The indices considered included population, gross national product and transport indices(26)

Sozen and Arcakliglu (2007) used neural network model to forecast energy consumption in Turkey. Three different models were used for network training. Based on this study, they concluded that neural network is more appropriate for forecasting net energy consumption using population and economical induces(27).

Kazemi et al. (2012) have proposed Gray Markov chain model for forecasting energy demand of transport sector in Iran. The forecasted results have then been compared with those of Gray and regression models. In addition, Gray Markov chain model has been used to forecast the energy demand of transport sector in Iran up to 1400(28).

Azadeh et al. (2008) have proposed a neural network model for forecasting annual electricity consumption of high energy consuming industrial sectors in Iran such as chemical, base metals, minerals and non-metals. Regression models were found to be not very precise given the variations in energy consumption. Therefore, a multi-layer perceptron neural network was used for forecasting and ultimately a model based on neural network and ANOVA was proposed for forecasting long term electricity consumption in high energy consuming industries(29).

Ekonomou (2010) proposed the multi-layer perceptron neural network for forecasting long term energy consumption in Greece and compared it with regression model. The results of then comparison indicated the efficiency of multi-layer perceptron neural network(30).

Limanondet al. (2011) used feed forward neural network model in linear regression for forecasting energy consumption in the next 20 years in the transport sector in Thailand and forecasted the energy consumption for 2010-2030. The input variables considered included population, number of vehicles and gross domestic product(31).

Neural network model was used by Geem (2011) to forecast energy consumption in the transport sector in South Korea. Variables such as gross domestic product, population, petroleum price, number of vehicles and the number of trips were considered as input variables. Neural network model was determined as the more efficient model(32).

Bastani (2012) used possible techniques and scoring models to forecast the production of greenhouse gases and fuel consumption under uncertainty conditions in the United States. The research indicated the probability distribution of greenhouse gas emissions and fuel consumption up to 2050. Parameters, which are major causes of variations in the production of greenhouse gases and fuel consumption, were identified and categorized by the analysis of uncertainty in this work. Some of the parameters considered by the authors were the kilometers traveled in a year and car sales(33).

3. Fuzzy linear regression

5.1. Fuzzy linear regression model

Fuzzy linear regression is a generalization of conventional linear regression, which is used for calculation of the functional relationship between dependent and independent variables in a fuzzy medium. This model has been used for various applications including forecasting energy. The inputs and outputs of the model are non-fuzzy data. The base model is considered as a linear function:

$$\tilde{y} = f(x, \tilde{A}) = \widetilde{A_0} + \widetilde{A_1}x_1 + \widetilde{A_2}x_2 + \ \ldots + \widetilde{A_n}x_n \qquad (1)$$

Where $\widetilde{A_i}(i = 0,1,2,\ldots,n)$ are fuzzy coefficients in (p_i, c_i) form such that $p_i$and $c_i$are the middle and range parameters. The value of the middle parameter indicates the extent of fuzziness. The membership function corresponding to$\widetilde{A_i}$ fuzzy coefficient is shown in the Figure 1. Equation 1 can be written as:

$$\tilde{y} = (p_0, c_0) + (p_1, c_1)x_1 + (p_2, c_2)x_2 + \ \ldots + (p_n, c_n)x_n \qquad (2)$$

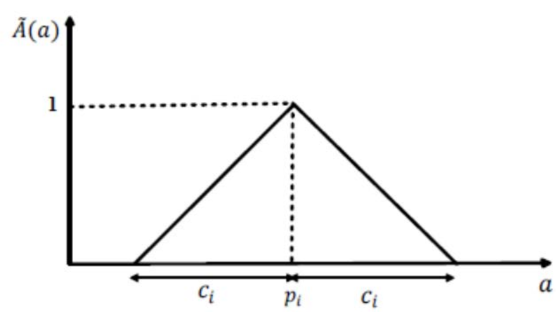

Figure 1: Membership function of $\widetilde{A_i}$fuzzy coefficient

The following membership function is obtained using the expansion principle (11):

$$\tilde{y} = \begin{cases} \max(\min_i \{\tilde{A}_i(a_i)\}) & \{a_i / y = f(x, a_i)\} \neq \phi \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Equations 4 are obtained based on Equations 2 and 3:

$$\tilde{y}(y)=\begin{cases}1-\dfrac{\left|y-\sum_{i=1}^{n}p_i x_i\right|}{\sum_{i=1}^{n}c_i\left|x_i\right|} & x_i\neq 0\\ 1 & x_i=0,\ y=0\\ 0 & x_i=0,\ y\neq 0\end{cases} \qquad \textbf{(4)}$$

Equation 2 can also be shown as follows:

$$\tilde{y}_i=(p_0,c_0)+(p_1,c_1)x_{1j}+(p_2,c_2)x_{2j}+\ \ldots+(p_n,c_n)x_{nj} \qquad \textbf{(5)}$$

j=1,2,…..m

Where m is the number of observations. The objective of regression model is the determination of optimal values of the parameters such that observation y_i with the degree of membership of at least *h* belongs to $\tilde{y}(y)$. Therefore, we have (13):

$$\tilde{y}(y)\geq h \qquad j=1,2,\ldots,m \qquad \textbf{(6)}$$

The degree of membership (h) is determined by the user. Figure 2 shows fuzzy output membership function. The fuzzy output must fall between A and B values.

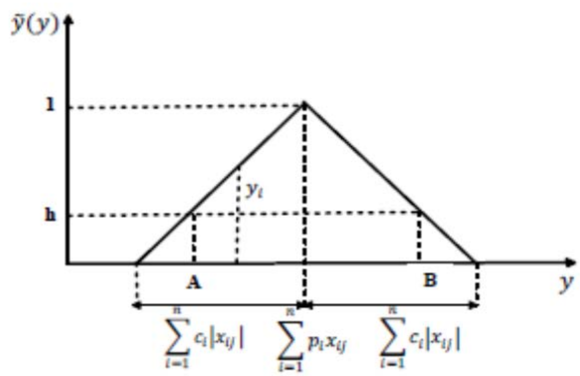

Figure 2: Fuzzy output membership function of regression model

Substitution of equation 4 in equation 6 gives equation 7:

$$\begin{aligned} y_j &\geq p_0+\sum_{i=1}^{n}p_i x_{ij}-(1-h)\Big(c_0+\sum_{i=1}^{n}c_i\left|x_{ij}\right|\Big), \qquad j=1,2,\ldots,m\\ y_j &\leq p_0+\sum_{i=1}^{n}p_i x_{ij}+(1-h)\Big(c_0+\sum_{i=1}^{n}c_i\left|x_{ij}\right|\Big), \qquad j=1,2,\ldots,m \end{aligned} \qquad \textbf{(7)}$$

5.2. Model Evaluation

One of the criteria for evaluation of linear regression models with fuzzy coefficient is based on the index of confidence (IC), defined as follows:

$$IC = 1 - \frac{SSE}{SST} \tag{8}$$

Where SSE and SST are the sum of squared errors and total squares of fuzzy linear regression, respectively which are shown in equation 9 (14).

$$SSE = 2\sum_{i=1}^{m} (y_j - \hat{y}_i^c)^2$$
$$SST = \sum_{i=1}^{m} (y_j - \hat{y}_i^L)^2 + \sum_{i=1}^{m} (y_j^R - y_j)^2 \tag{9}$$

5.3. *Selection of optimized model*

In linear regression with fuzzy coefficients, different models with Kurtosis coefficients may be considered and ultimately a model with the smallest IC is selected as the optimized model. Of course, an appropriate level of validity should also be considered.

It is necessary to improve the central coefficients here since fuzzy uncertainty in symmetrical form does not greatly influence the numerical center. The variation of h level does not affect the central value (center of coefficients) and only influences the widths of fuzzy numbers (model coefficients). Therefore it is necessary to use a method to improve middle values. As it is known, model coefficients in fact define model slope and the fuzzy width is indeed equivalent to the standard error in conventional model. However, a function can be defined as the error function, which consists of the set of absolute values of model errors so as not to contradict the resulting degree of constraints which is shown in equation 10.

$$MinZ_e = \sum_{j=1}^{n} \left| Y_j - (C_0 + \sum_{i=1}^{m} C_i X_{ji}) \right|$$
$$(1-h)S_0 + (1-h)\sum_{i=1}^{n} (S_i X_{ji}) + C_0 + \sum_{i=1}^{m} (C_i X_{ji}) \geq y_j$$
$$(1-h)S_0 + (1-h)\sum_{i=1}^{n} (S_i X_{ji}) - C_0 - \sum_{i=1}^{m} (C_i X_{ji}) \geq -y_j \tag{10}$$

Considering the above function and the level of fuzzy width compatible with IC, it is well optimized using genetic algorithm in order to obtain better central coefficients compared with the previous fuzzy model resulting from conventional regression. After finding the unfuzzified fuzzy regression model using the genetic algorithm, the error from this model is evaluated and the results are compared with those of conventional regression.

4. Determination of dependent and independent variables

Dependent and independent variables corresponding to the demands for each of the petroleum products including gasoline, kerosene, fuel oil, gasoil and LPG and the method of their determination will be discussed in this section. As it was previously stated, the inputs of different models have been selected according to the researches carried out and the experts' comments. Each of the forecasting models for petroleum products include independent variables as output, forecasting function and the demand for petroleum products as the output. Linear, Cobb-Douglas and fuzzy linear regression models have been used as forecasting function in this research. The models have then been compared and the most appropriate model has been selected as the forecasting function. In addition, the data for the years 1996-2016 has been used for designing the proper forecasting model. 1996-2011 data have been used for the model evaluation and those for the years 2012-2016 have been applied in testing the model. After the most proper forecasting model is selected, it is necessary to assess the demand of petroleum products in 2017-2036. This requires the estimation of input variables and/or independent variables each of which have been assessed using an appropriate function. The forecasting model of each of the petroleum products follows. The following figure shows the forecasting model for gasoline, gasoil, fuel oil, kerosene and LPG demand.

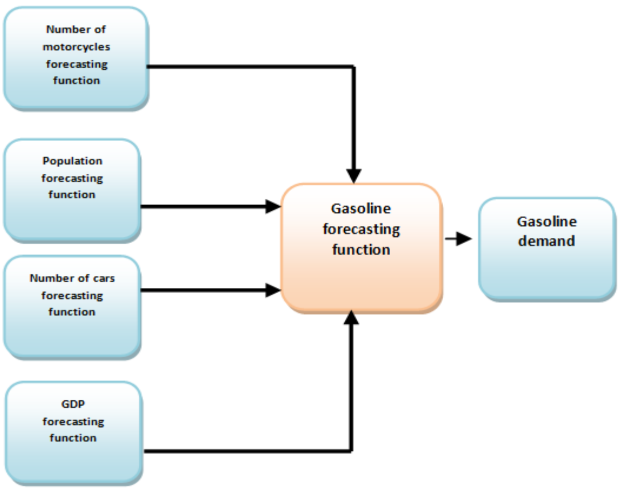

Figure 3: Forecasting model for gasoline demand

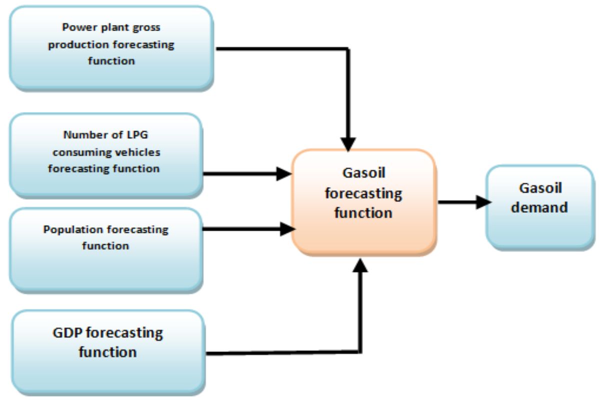

Figure 4: Forecasting model for gasoil demand

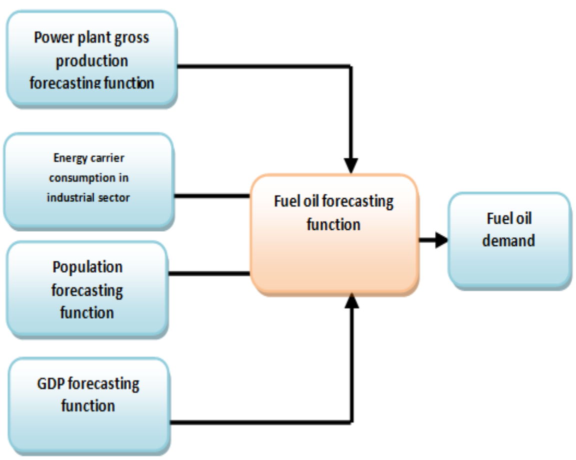

Figure 5: Forecasting model for fuel oil demand

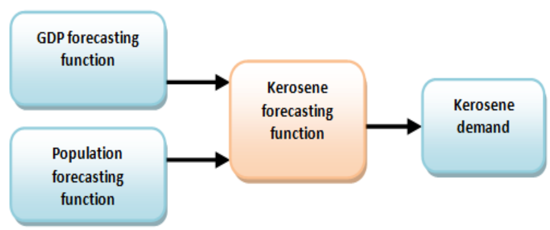

Figure 6: Forecasting model for kerosene demand

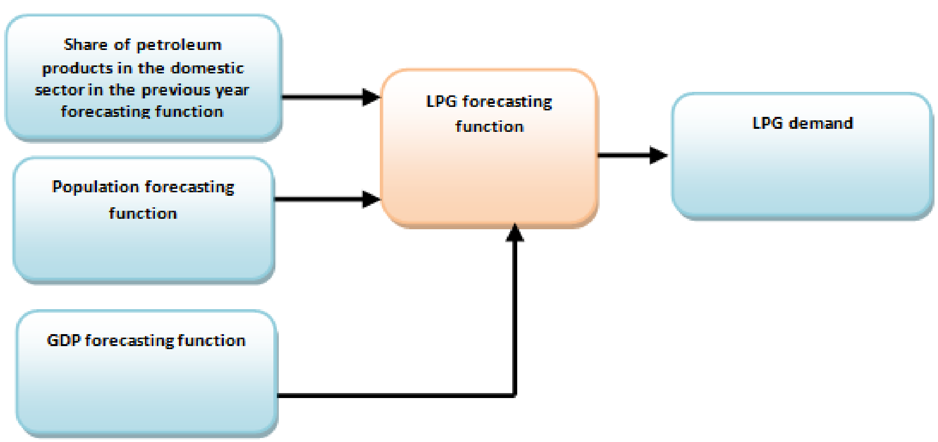

Figure 7: Forecasting model for LPG demand

5. Evaluation of forecasting function

The forecasting function for the demand of each of the petroleum products by linear regression, regression using Cobb-Douglas function and fuzzy linear regression is evaluated here. Data for the years 1996-2011 are used for evaluation, as previously stated. The forecasting functions evaluated will then be compared using the data for the years 2012-2016 and the most appropriate function will be selected considering the evaluation criterion. The evaluated functions for gasoline are as follows:

5.1. Linear regression

Linear regression using Minitab software has been used for evaluation of gasoline using the data for the years 1996-2011 as follows:

$$y = -15465046 + 1.57\, x_1 + 0.337\, x_2 + 50.0\, x_3 + 212\, x_4 \tag{11}$$

$$S = 402496 \quad \text{R-Sq} = 99.6\% \quad \text{R-Sq(adj)} = 99.5\% \tag{12}$$

As observed, the adjusted coefficient of determination is 99.5%, which is appropriate.

5.2. Regression using Cobb-Douglas function

The regression line by Cobb-Douglas function using Minitab software for evaluation of gasoline has been evaluated using the data for the years 1996-2011 as follows:

$$y = \beta_0 {x_1}^{\beta_1} {x_2}^{\beta_2} {x_3}^{\beta_3} {x_4}^{\beta_4} {x_5}^{\beta_5} {x_6}^{\beta_6} {x_7}^{\beta_7} {x_8}^{\beta_8} {x_9}^{\beta_9} + \varepsilon \tag{13}$$

Considering the non-linearity of equation 13 based on regression model parameters, the natural logarithm of both sides of the relationship is taken:

$$\begin{aligned} ln\, y \; = ln\, \beta_0 \; + \beta_1 ln\, x_1 \; + \beta_2 ln\, x_2 \; + \beta_3 ln\, x_3 \; + \beta_4 ln\, x_4 \; + \beta_5 ln\, x_5 \; + \beta_6 ln\, x_6 \\ + \beta_7 ln\, x_7 \; + \beta_8 ln\, x_8 \; + \beta_9 ln\, x_9 \end{aligned} \tag{14}$$

Minitab software output for this model is as follows:

$$y = -24.7 - 0.0164x_1 - 0.026x_2 + 0.734x_3 - 1.08x_4 + 2.94x_5 \tag{15}$$

$$S = 0.0238853 \quad \text{R-Sq} = 99.6\% \quad \text{R-Sq(adj)} = 99.4\%$$

As observed, the adjusted coefficient of determination is 99.4%, which is appropriate. The analysis of variance (ANOVA) table is given below:

5.3. Regression using fuzzy linear regression

Using the data for the years 1996-2011, the model proposed in equation 10 must be written for the corresponding problem for gasoline using fuzzy regression model. Lingo 11.0 software has been used in this work to solve the model. The optimized coefficients of the fuzzy regression model obtained by solving the non-linear programming model are shown in Table 1.

Table 1: Optimized coefficients of the fuzzy regression model

(p_0, c_0)	(p_1, c_1)	(p_2, c_2)	(p_3, c_3)	(p_4, c_4)
(0, 9500001)	**(3.124652, 1.258086)**	**(0, 1.396496)**	**(44.14147, 1.239543)**	(0, 1.234510)

6. Selecting the most appropriated the predictive function

The forecasting functions were evaluated in the previous section using three regression functions including linear regression, regression using Cobb-Douglas function and fuzzy linear regression and the data for the years 1996-2011. The most appropriate forecasting function must now be selected. The data for the years 2012-2016 are used as test data for the comparison of the three models. Mean absolute percentage error (MAPE) evaluation criteria is used for the comparison of the three models. Generally, there are four error evaluation methods including mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and mean absolute percentage error (MAPE). MAPE method has been used in this work since this is a conventional method for evaluation of relative error. All methods, except for MAPE, have a measurement unit. MAPE is a very appropriate method for the evaluation of errors when the model input variables have different measurement units or scales. MAPE is calculated as follows:

$$MAPE = \frac{\sum_{t=1}^{n}\left(\left|x_t - x^t / x_t\right|\right)}{n} \tag{16}$$

The results of forecasting gasoline consumption during 2012-2016 using independent variables in the years 2012-2016 by the three regression functions are shown in Table 2.

Table 2: The results of forecasting gasoline consumption during 2012-2016

Year	Real consumption	Forecasting consumption (fuzzy regression)	Forecasting consumption (Cobb-Douglas models)	Forecasting consumption (linear regression)
2012	24168723	24456304.1	28596980	**27979019**
2013	24496432	25009557.5	29518480	**28846849**
2014	23619352	26165173.1	30999199	**30369629**
2015	22365183	27299799.2	33354771	**32263611**
2016	21879197	29095911.5	35726073	**34395073**
MAPE		0.138223	0.3649	**0.327134**

It should be noted that the MAPE criterion for fuzzy regression is considerably lower (about 57%) compared with linear regression and Cobb-Douglas models, which indicates the better performance of fuzzy regression model.

Analysis of variance (ANOVA) is used for more analysis of the results. ANOVA is an appropriate method for testing the equality of variances (34). The test must be designed such that deviations from external sources are systematically controlled. Time is a common source of deviations, which can be systematically controlled by building blocks. Therefore, blocked design of ANOVA may be used. The hypothesis testing is as follows:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 H_0: \mu_i \neq \mu_j \ i, j = 1,2,3,4, i \neq j \tag{17}$$

Where μ_1, μ_2, μ_3 and μ_4 are the average forecasted values by regression using Cobb-Douglas function, average actual values, average forecasted values by linear regression and average forecasted values by fuzzy linear regression, investigated in α significance level. If the null hypothesis is accepted, the model with MAPE will be accepted. Otherwise, Tukey test will be used to compare the four models and select the most appropriate one.

The analysis of variance is carried out based on block design. As indicated in the previous section, the data for the years 1996-2011 will be used for the evaluation of model parameters and those for the years 2012-2016 will be used for testing the model. According to ANOVA results (Table 3), it is observed that the null hypothesis is rejected at a significance level of 5%. In other words, this result may indicate that one of the average values is different from the others. Considering the ANOVA results, more analysis is required to determine the most appropriate forecasting model. Therefore, Tukey test will be used.

Table 3: ANOVA results at a significance level of 5%

Tests of Between-Subjects Effects Dependent Variable: VAR00003					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	266923058478536.000(a)	7	38131865496933.720	11.301	.000
Intercept	15712240082696290.000	1	15712240082696290.000	4656.512	.000
Treat	227575083772724.800	3	75858361257574.900	22.482	.000
Block	39347974705811.140	4	9836993676452.780	2.915	.067
Error	40491010231315.520	12	3374250852609.627		
Total	16019654151406150.000	20			
Corrected Total	307414068709851.600	19			
R Squared = .868 (Adjusted R Squared = .791)					

Tukey (1953) proposed a multiple comparison method based on Student's test range. The method requires qα (α, f) value to determine the critical value for all pairwise comparisons without considering the number of averages placed in one group (34, 35). Therefore, the value of Tukey test states whether the two averages are considerably different. The two averages are equal if their absolute value, which is affected by their sample size, is considerably greater than the value obtained from equation 18.

$$T_{\alpha} = q_{\alpha}(\alpha, f) S_{\bar{y}_i} \tag{18}$$

q_{α} (α, f) is the range of Tukey table and $s\underline{y}_i$ equals $\sqrt{\frac{MS_E}{n}}$.

The Tukey test is carried out considering the results of ANOVA which is shown in figure 8:

```
One-way ANOVA: LN, REAL, REG, FUZZ
Source  DF           SS           MS      F      P
Factor   3  2.27575E+14  7.58584E+13  15.20  0.000
Error   16  7.98390E+13  4.98994E+12
Total   19  3.07414E+14

S = 2233817   R-Sq = 74.03%   R-Sq(adj) = 69.16%

                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level  N      Mean    StDev  ---------+---------+---------+---------+
LN     5  31639100  2907227                         (-----*-----)
REAL   5  23305777  1138037  (-----*-----)
REG    5  30770836  2599058                       (-----*-----)
FUZZ   5  26399370  1859448          (-----*-----)
                             ---------+---------+---------+---------+
                              24500000  28000000  31500000  35000000
Pooled StDev = 2233817

Grouping Information Using Tukey Method
      N      Mean  Grouping
LN    5  31639100    A
REG   5  30770836    A
FUZZ  5  26399370    B
REAL  5  23305777    B

Tukey 95% Simultaneous Confidence Intervals
Individual confidence level = 98.87%
```

Figure 8: Forecasting model for LPG demand

Considering the results obtained, it is observed that only the averages of fuzzy regression population and actual data are equal and the averages of other pairwise comparisons including actual data and regression using Cobb-Douglas function are not equal to the average of actual data. Therefore, the most appropriate function for forecasting petroleum demand is the fuzzy regression function, according to the results.

7. Forecasting demands for petroleum products

Fuzzy regression model was selected as the most appropriate model for forecasting gasoline demand in the previous section according to MAPE criterion, ANOVA and Tukey test. Therefore, this model is used to forecast gasoline demand in 2017-2036. The point to ponder here is that the values of independent variables during 2017-2036 are required for forecasting. Thus, the values of independent variables in 2017-2036 are forecasted considering their trend and the demand is then forecasted using fuzzy regression function.For this purpose, scatter plots of independent variables vs. time are first plotted, as shown in figure 9.

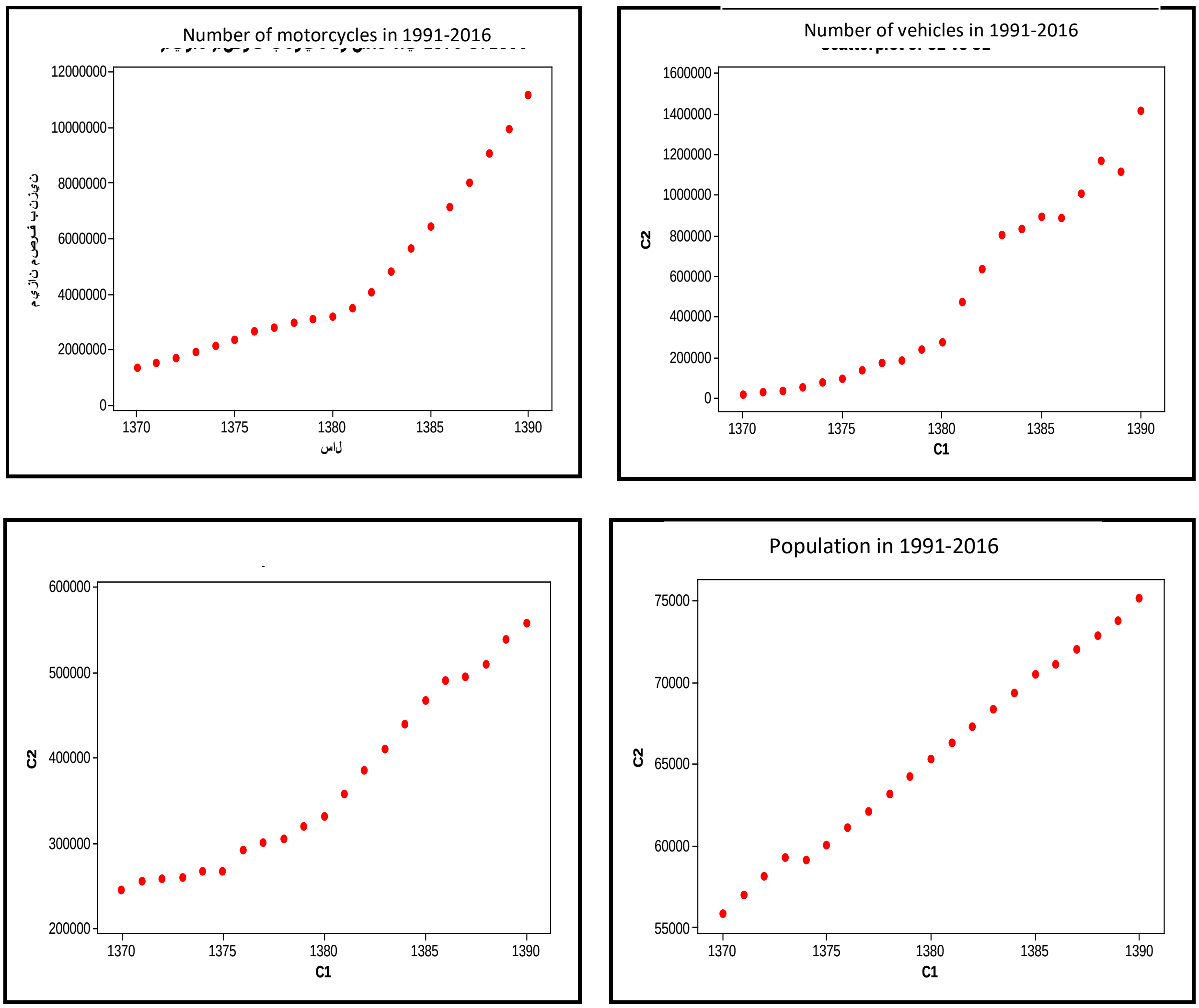

Figure 9: Scatter plots of independent variables vs. time

A non-linear trend is observed between dependent and independent variables considering the scatter plots of the number of motorcycles, vehicles and gross domestic product vs. time. According to the trend type, a quadratic function is used for the relationship between dependent and independent variables, which is shown in equations 19-21.

C2 = 5.44E+10 - 79235341 X + 28872 X^2 **(19)**

Y= 5552879000 - 8118003X+2967*X^2 **(20)**

Y=1.14E+9-1671709*X+611.7*(X^2) **(21)**

The population in the corresponding years is forecasted using a linear function (equation 22) considering the trend in the population.

Y=- 1257967 + 958.9*X **(22)**

Gasoline demand in 2017-2036 is forecasted using fuzzy regression function considering the estimated values for dependent variables, as shown in the following table 3. According to the table 3, the lower and upper bounds are in fact the forecasted confidence interval.

Table 3: Forecasted gasoline demand over time

Year	Upper bound	Central value	Lower bound	Year	Upper bound	Central value	Lower bound
2017	33766388	31106712.3	28447037	**2027**	57442415	52249699.2	**47056983**
2018	35767150	32894561.4	30021973	**2028**	60258379	54762991.9	**49267605**
2019	37849432	34754954.8	31660478	**2029**	63155862	57348828.9	**51541795**
2020	40013234	36687892.6	33362551	**2030**	66134866	60007210.3	**53879554**
2021	42258557	38693374.8	35128193	**2031**	69195390	62738136.1	**56280882**
2022	44585400	40771401.3	36957403	**2032**	72337434	65541606.2	**58745778**
2023	46993762	42921972.1	38850182	**2033**	75560999	68417620.6	**61274243**
2024	49483645	45145087.4	40806529	**2034**	78866083	71366179.5	**63866276**
2025	52055049	47440747	42826445	**2035**	82252688	74387282.6	**66521878**
2026	54707972	49808950.9	44909930	**2036**	85720812	77480930.2	**69241048**

In addition, the forecasted central value and upper and lower bounds for 2016-2037 are shown in Figure 10.

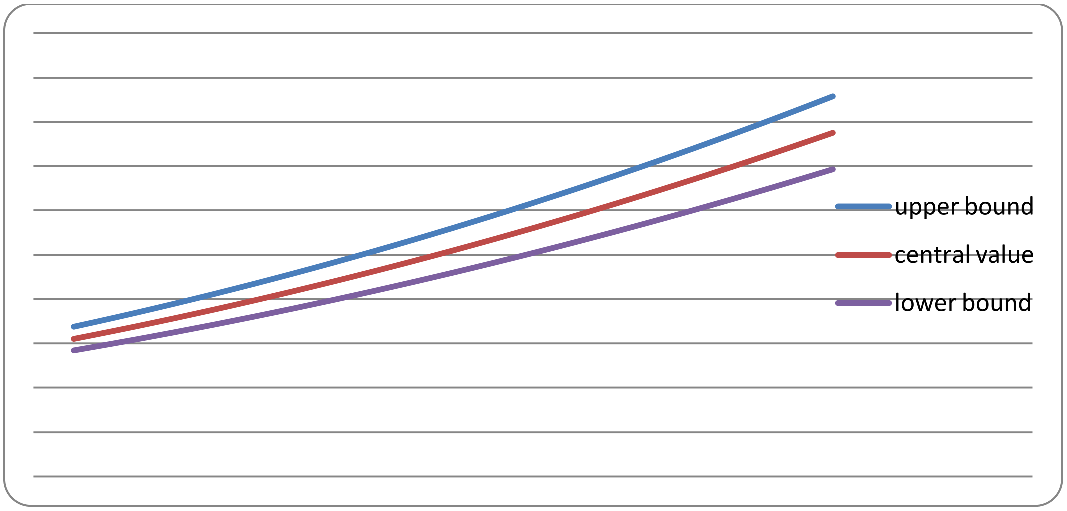

Figure 10: Gasoline forecasted demand diagram

The comparison of the three regression functions has also been carried out for the other petroleum products (gasoil, fuel oil, kerosene and LPG) and the fuzzy linear regression function has been selected. Fuzzy linear relationships for gasoil, fuel oil, kerosene and LPG are shown in tables 4-7:

Table 4: Optimized coefficients for gasoil fuzzy regression model

(p0, c0)	(p1, c1)	(p2, c2)	(p3, c3)	(p4, c4)
(0, 5213848)	(25.98756, 0)	(0, 0)	(320.2, 0)	**(156.27, 0)**

Table 5: Optimized coefficients for fuel oil fuzzy regression model

(p0, c0)	(p1, c1)	(p2, c2)	(p3, c3)	(p4, c4)
(13883220, 1240517)	(0.006628693, 0)	(0, 0)	(13.25883, 0)	**(0, 3390)**

Table 6: Optimized coefficients for kerosene fuzzy regression model

(p0, c0)	(p1, c1)	(p2, c2)
(8610596.38, 945676)	(0, 2.234568)	**(0, 0)**

Table 7: Optimized coefficients for LPG fuzzy regression model

(p0, c0)	(p1, c1)	(p2, c2)	(p3, c3)
(24689076, 837490)	(2.873156, 1.234568)	(0, 1.23456)	**(0, 1.234568)**

The values forecasted for the demands of gasoil, fuel oil, kerosene and LPG in 2017-2036 are shown in Tables 8, 9, 10 and 11, respectively.

Table 8: Forecasted gasoil demand over time

Year	Upper bound	Central value	Lower bound	Year	Upper bound	Central value	Lower bound
2017	43799271	38585423	33371575.5	**2027**	59735170	54521322	**49307474.5**
2018	45202319	39988471	34774622.8	**2028**	61561646	56347798	**51133949.7**
2019	46647709	41433861	36220013	**2029**	63430464	58216616	**53002767.7**
2020	48135442	42921594	37707745.9	**2030**	65341625	60127777	**54913928.6**
2021	49665518	44451670	39237821.6	**2031**	67295128	62081280	**56867432.1**
2022	51237936	46024088	40810240.1	**2032**	69290975	64077127	**58863278.5**
2023	52852697	47638849	42425001.4	**2033**	71329164	66115316	**60901467.7**
2024	54509802	49295954	44082105.5	**2034**	73409696	68195848	**62981999.7**
2025	56209248	50995400	45781552.4	**2035**	75532570	70318722	**65104874.4**
2026	57951038	52737190	47523342	**2036**	77697788	72483940	**67270091.9**

Table 9: Forecasted fuel oil demand over time

Year	Upper bound	Central value	Lower bound	Year	Lower bound	Central value	Upper bound
2017	17200768.39	14890804	12580840	**2027**	11570546	15019099	**18467651.08**
2018	17303470.58	14903619	12503768	**2028**	11440236	15031946	**18623655.67**
2019	17411503.01	14916437	12421371	**2029**	11304603	15044797	**18784990.49**
2020	17524865.68	14929258	12333651	**2030**	11163645	15057650	**18951655.56**
2021	17643558.59	14942083	12240607	**2031**	11017364	15070508	**19123650.86**
2022	17767581.74	14954911	12142240	**2032**	10865760	15083368	**19300976.41**
2023	17896935.13	14967742	12038548	**2033**	10708831	15096232	**19483632.19**
2024	18031618.76	14980576	11929533	**2034**	10546579	15109099	**19671618.21**
2025	18171632.63	14993414	11815195	**2035**	10379003	15121969	**19864934.48**
2026	18316976.73	15006255	11695532	**2036**	10206104	15134842	**20063580.98**

Table 10: Forecasted kerosene demand over time

Year	Upper bound	Central value	Lower bound	Year	Upper bound	Central value	Lower bound
2017	10889168	8610597	6332027	**2027**	11697140	8610607	**5524075**
2018	10957663	8610598	6263534	**2028**	11792973	8610608	**5428244**
2019	11028892	8610599	6192307	**2029**	11891539	8610609	**5329680**
2020	11102855	8610600	6118346	**2030**	11992840	8610610	**5228381**
2021	11179551	8610601	6041651	**2031**	12096874	8610611	**5124349**
2022	11258982	8610602	5962223	**2032**	12203642	8610612	**5017583**
2023	11341146	8610603	5880061	**2033**	12313144	8610613	**4908083**
2024	11426044	8610604	5795165	**2034**	12425379	8610614	**4795850**
2025	11513675	8610605	5707536	**2035**	12540348	8610615	**4680882**
2026	11604040	8610606	5617172	**2036**	12658052	8610616	**4563181**

Table 11: Forecasted LPG demand over time

Year	Upper bound	Central value	Lower bound	Year	Upper bound	Central value	Lower bound
2017	28070438.88	26402881.09	25471729.16	**2027**	29567520.14	27441739.59	**26498752.60**
2018	28197531.49	26490949.35	25558615.18	**2028**	29744868.23	27564958.05	**26620787.22**
2019	28329649.45	26582532.64	25649016.26	**2029**	29927241.71	27691691.52	**26746336.87**
2020	28466794.43	26677630.94	25742930.73	**2030**	30114640.58	27821940.02	**26875401.54**
2021	28608964.79	26776244.26	25840360.23	**2031**	30307064.84	27955703.53	**27007981.22**
2022	28756160.54	26878372.60	25941304.74	**2032**	30504514.49	28092982.06	**27144075.93**
2023	28908381.68	26984015.96	26045764.27	**2033**	30706989.52	28233775.61	**27283685.65**
2024	29065628.21	27093174.34	26153738.83	**2034**	30914489.95	28378084.18	**27426810.39**
2025	29227900.13	27205847.74	26265228.40	**2035**	31127015.77	28525907.77	**27573450.16**
2026	28070438.88	26402881.09	25471729.16	**2036**	31344566.97	28677246.38	**27723604.94**

7. Conclusion

Given the dual function of the energy sector in energy supply and foreign exchange income, energy sector is considered the infrastructure for development and has had a fundamental part in the social/economical parts of the country. It is of great importance to address the consumption of petroleum products in different sections, which constitutes 46.1% of the entire energy carriers. The forecasting of five main petroleum products including gasoline, gasoil, fuel oil, kerosene and LPG has been performed in this work given the significance of petroleum product sector in supplying the energy demands of the country and necessity of through and precise planning with regards to their utilization. Linear regression, regression using Cobb-Douglas function and fuzzy regression functions have been used for forecasting. The results of the three functions have then been compared using block analysis of variance and Tukey test. The results have shown the effectiveness of fuzzy linear regression model. Finally, the demands for the five petroleum products have been forecasted in 2017-2036 using fuzzy linear regression function. According to the results of forecasting, gasoline and gasoil demands are observed to considerably increase in the next twenty years, indicating that these products are strategic and their demand requites a comprehensive, long term planning. The prices of petroleum products have been considered effective factors on forecasting in this work. The actual prices of petroleum products and price liberalization since 2010 were issues, which needed more extensive analysis and investigations to take account for other factors in petroleum products and are suggested as future research subjects. One of the other subjects to further research into is the effect of alternative fuel products.

References

1. Jafarzadeh H, Gholami S, Bashirzadeh R. A new effective algorithm for on-line robot motion planning. Decision Science Letters. 2014;3(1):121-30.
2. Snyder LV, Daskin MS. A random-key genetic algorithm for the generalized traveling salesman problem. European journal of operational research. 2006 Oct 1;174(1):38-53.
3. Kumar U, Jain V. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. Energy. 2010;35(4):1709-16.
4. Pokharel S. An econometric analysis of energy consumption in Nepal. Energy Policy. 2007;35(1):350-61.
5. Kucukali S, Baris K. Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach. Energy Policy. 2010;38(5):2438-45.
6. Marvuglia A, Messineo A. Using recurrent artificial neural networks to forecast household electricity consumption. Energy Procedia. 2012;14:45-55.
7. Forouzanfar M, Doustmohammadi A, Hasanzadeh S. Transport energy demand forecast using multi-level genetic programming. Applied Energy. 2012;91(1):496-503.
8. Kıran MS, Özceylan E, Gündüz M, Paksoy T. A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey. Energy conversion and management. 2012;53(1):75-83.
9. Ünler A. Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025. Energy Policy. 2008;36(6):1937-44.
10. Moradi Nasab N, Amin-Naseri M, Behbahani TJ, Nakhai Kamal Abadi I. An integrated economic model of fossil fuel energy planning for government and private sectors. Energy Sources, Part B: Economics, Planning, and Policy. 2016;11(7):651-64.

11. Nasab NM, Amin-Naseri M. Designing an integrated model for a multi-period, multi-echelon and multi-product petroleum supply chain. Energy. 2016;114:708-33.

12. H. Tanaka SU, K. Asai. Linear regression analysis with fuzzy model. IEEE Trans Systems Man Cybernet. 1982;12

13. Heshmaty B, Kandel A. Fuzzy linear regression and its applications to forecasting in uncertain environment. Fuzzy sets and systems. 1985;15(2):159-91.

14. Wang H-F, Tsaur R-C. Resolution of fuzzy regression model. European Journal of Operational Research. 2000;126(3):637-50.

15. Chang P-T, Lee ES, Konz SA. Applying fuzzy linear regression to VDT legibility. Fuzzy Sets and Systems. 1996;80(2):197-204.

16. Lai Y-J, Chang SI. A fuzzy approach for multiresponse optimization: An off-line quality engineering problem. Fuzzy Sets and Systems. 1994;63(2):117-29.

17. Al-Kandari A, Soliman S, El-Hawary M. Fuzzy short-term electric load forecasting. International Journal of Electrical Power & Energy Systems. 2004;26(2):111-22.

18. Taghizadeh M, Shakouri H, Menhaj M, Mehregan M, Kazemi A, editors. Design of a multi-level fuzzy linear regression model for forecasting transport energy demand: A case study of Iran. Computers & Industrial Engineering, 2009 CIE 2009 International Conference on; 2009: IEEE.

19. Nadimi R. A novel fuzzy linear regression model based on a non-equality possibility index and optimum uncertainty. Applied Soft Computing. 2009;9(2):590-8.

20. Azadeh A, Saberi M, Seraj O. An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran. Energy. 2010;35(6):2351-66.

21. Al-Omishy HK. Prediction of the fuel consumption of petrol and diesel vehicles from a computer-simulation model. Energy. 1989;14(9):563-70.

22. Nizami SJ, Al-Garni AZ. Forecasting electric energy consumption using neural networks. Energy policy. 1995;23(12):1097-104.

23. Kalogirou SA, Bojic M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy. 2000;25(5):479-91.

24. Kermanshahi B, Iwamiya H. Up to year 2020 load forecasting using neural nets. International Journal of Electrical Power & Energy Systems. 2002;24(9):789-97.

25. Nasr G, Badr E, Joun C. Backpropagation neural networks for modeling gasoline consumption. Energy Conversion and Management. 2003;44(6):893-905.

26. Murat YS, Ceylan H. Use of artificial neural networks for transport energy demand modeling. Energy policy. 2006;34(17):3165-72.

27. Sözen A, Arcaklioglu E. Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey. Energy policy. 2007;35(10):4981-92.

28. Kazemi AS, H. Menhaj, M.B. Mehregan, M.R and Neshat, N. Commercial Energy Demand Forecast: Iran Case Study. 3rd International Conference on Information and Financial Engineering 2011.

29. Azadeh A, Ghaderi S, Sohrabkhani S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. Energy Conversion and Management. 2008;49(8):2272-

8.

30. Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks. Energy. 2010;35(2):512-7.

31. Limanond T, Jomnonkwao S, Srikaew A. Projection of future transport energy demand of Thailand. Energy policy. 2011;39(5):2754-63.

32. Geem ZW. Transport energy demand modeling of South Korea using artificial neural network. Energy policy. 2011;39(8):4644-50.

33. Bastani P, Heywood JB, Hope C. Us cafe standards. MIT. Edu; 2012.

34. Design DM. Analysis of Experiments. New York: Tohn Wiley and Sons. 2001;1:976.

35. Tukey JW. The Problem of Multiple Comparisons: Introduction and Parts A, B, and C: B [Princeton University]; 1953.