Scientific Research and Essays Vol. 6(23), pp. 5072-5081, 16 October, 2011
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE11.1317
ISSN 1992-2248 © 2011 Academic Journals

Full Length Research Paper

Design of a multi-level fuzzy linear regression model for forecasting: A case study of Iran

M. R. Taghizadeh[1,2]*, H. Shakouri G.[3], E. Asgharizadeh[1] and M. Sakawa[4]

[1]Department of Industrial Management, Faculty of Management, University of Tehran, Tehran, Iran.
[2]Frontier Research Center, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan.
[3]Department of Industrial Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran.
[4]Department of System Cybernetics, Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Japan.

Accepted 29 August, 2011

Linear regression has been widely used for many years to forecast various socio-economic variables in marketing, management, sales, energy and so on. Demand and price are commonly estimated by means of regression models. However, where there is high uncertainty in the model, fuzzy regressions are applied. In this paper, a fuzzy-based approach is applied for transport energy demand forecasting using socio-economic and transport related indicators. The model is based on Gross Domestic Product (GDP), population and the number of vehicles as three inputs to the main model. Energy data from 1993 to 2005 are used to estimate each of the three inputs. By three individual fuzzy linear regression (FLR) models, a multi-level FLR model is designed. The input variables are transport energy demand in the last year, the number of vehicles, population and ratio of GDP over population. The output variable is energy demand of the transportation sector in Iran. The inputs to the ending level are obtained as outputs of the starting levels. The estimation fuzzy problem for the model is formulated as a linear optimization problem. Comparison of the model predictions with data of the testing period shows validity of the proposed model. Furthermore, having obtained the fuzzy parameters, the transport energy demand is predicted for 2006 to 2020. It is noticeable that, without any price shock or efficiency improvement in the transportation sector, the energy consumption may reach to a threatening level of about 592 MBOE by 2020.

Key words: Forecasting, multi-level fuzzy linear regression, transport energy demand, case study.

INTRODUCTION

Energy demand has rapidly been increasing because of developments in the industrial, agricultural, transportation, commercial and housing sectors. Population and GDP growth causes energy demand growth in all sectors. Moreover, the fast growth of the GDP leads to an increase in the number of vehicles for the individuals and consequently an increase in energy demand in the transportation sector. Foreseeing this Iran expects a very large growth in energy demand in the future, especially for gasoline in the transportation sector.

*Corresponding author. E-mail: mrtaghizadeh@ut.ac.ir. Tel: +81-80-4558-1355. Fax: +98-21-8800-8990.

This sector takes one of the biggest shares in energy use in Iran for instance about 27% of total energy consumption. Therefore, it is very important to forecast transport energy demand. Trying to fit some phenomena that are vague in nature into precise mathematical models does not always work. Zadeh (1965) described this discrepancy between real situations and applied mathematics by saying, "classical mathematics was too precise to describe systems where a human element was involved". When human judgments are involved with a system or the data obtained from the system is scarce or insufficiently precise, the vagueness of such a system considered must be dealt with in the system modeling. The fuzzy regression model has been applied to modeling systems involving vague or imprecise phenomena.

Al-Kandari et al. (2004) have developed a fuzzy linear regression model for forecasting summer and winter season's related load. The estimation fuzzy problem for the model is turned out to a linear optimization problem, fuzzy linear regression. It has been found that using such fuzzy model, a reliable operation for the electric power system could be obtained. Soliman et al. (2003) have proposed a new technique for frequency and harmonic evaluation in power networks. This technique is based on fuzzy linear regression and uses the digitized voltage samples to estimate the frequency and harmonic contents of the voltage signal. Effects of sampling frequency, data window size, and the degree of fuzziness on the estimated parameters are investigated. Shakouri et al. (2008) have proposed a hybrid model by combining a fuzzy regression with a Takagi-Sugeno-Kang (TSK) model to study the effects of climate change in electricity consumption. Srivastava and Nema (2008) used fuzzy linear regression for forecasting the solid waste composition of Delhi, India over the period 2007 to 2024. The findings emphasized the importance of forecasting waste composition and the significance of waste segregation for efficient operation of various reuse-recycle, treatment and disposal facilities. Heshmaty and Kandel (1985) applied fuzzy linear regression to forecast the computer sales in the United States. Their entire fuzzification process is based on Armstrong's model (which deals with forecasting the international camera sales) regarding the camera market. Bell and Wang (1997) built fuzzy linear regression models to reveal the relationship of cumulative trauma disorder (CTD) risk factors, to predict the injuries, and to evaluate risk levels of individuals. Four fuzzy models were built corresponding to four risk categories, and a final fuzzy linear model was established using AHP pairwise comparisons.

From a methodological point of view, fuzzy linear regression models provide useful insight into risk factors-CTD relationship. Akkuş and Asilturk (2011) have modeled the average surface roughness values using fuzzy logic, artificial neural networks (ANN) and multi-regression equations. Input variables are cutting speed (V), feed rate (f) and depth of cut (a) while output variable is surface roughness (Ra). The predicted values of mean squared errors (MSE) were appplied to compare the three models. Results showed that the optimum predictive model is the fuzzy logic model. Yalpir and Özkan (2011) have evaluated residential real estates existing at the Konya, Turkey using fuzzy logic and multiple regression techniques. Input variables are area, age, floor condition, physical properties and location of the real-estate, and the market value of the estate was determined as the output variable. According to the results R^2 and root mean square error (RMSE) values were 0.85 and 19 for fuzzy inference systems (FIS), 0.64 and 30 for multiple regression analysis (MRA), respectively. Sarah et al. (2011) have presented a long term forecasting method by a combination of intelligent methods with the use of the past month rainfall in karoon basin and global meteorological signals such as southern oscillation index (SOI), north athletics oscillation (NAO), sea level pressure (SLP), sea surface temperature (SST) and 41 years historical data. This method is obtained by the combination of artificial neural network, fuzzy logic and wavelet functions and the long-term forecasts are done for periods of six months, one year and two years. As a result of the root mean squared error, predicting the two-year and annual periods is 6.22 and 7.11, respectively. But, the predicted six months shows 13.15. Linear regression model is designed for the transport energy demand forecasting using gross domestic product (GDP), population and the number of vehicles in Iran for the time span 2006 to 2020. The estimation fuzzy problem for the model is formulated as a linear optimization problem. The fuzzy parameters are estimated using the past history data from 1993 to 2005.

The results show that, without any price shock or efficiency improvement in the transportation sector, the energy consumption may reach to a threatening level of about 592 MBOE by 2020. The remainder of this paper is organized as follows: subsequently, we discuss the concept of fuzzy linear regression which is being used in Section 3 to build the forecasting models for the transport energy demand (GDP, population and the number of vehicles) in Iran. The framework of the forecasting will be discussed subsequently. Fuzzy models and obtained results are presented.

FUZZY LINEAR REGRESSION

Fuzzy linear regression was proposed by Tanaka et al. (1982). This method is widely applied to various applications including marketing, management and sales forecasting (Heshmaty and Kandel, 1985; Chang, 1997). In conventional regression techniques, the difference between the observed values and the values estimated from the model is assumed to be due to observational errors but in fuzzy regression, the difference between the observed and the estimated values is assumed to be due to the ambiguity inherently present in the system. In this section, a formulation for fuzzy linear regression estimation problem is presented. In this model, the outputs are non-fuzzy observations. Also, the inputs are non-fuzzy inputs. The base model is assumed to be a fuzzy linear function as follows:

$$\tilde{y} = f\left(x, \tilde{A}\right) = A_0 + \tilde{A}_1 X_1 + \tilde{A}_2 X_2 + \ldots + \tilde{A}_n X_n \tag{1}$$

Where A_i (I = 1, 2, …, n) are the fuzzy coefficients in the form of $\left(p_i, c_i, c_i\right)$ where p_i is the middle and c_i is the spread. The spread value denotes the fuzziness of the function. The membership function for the fuzzy

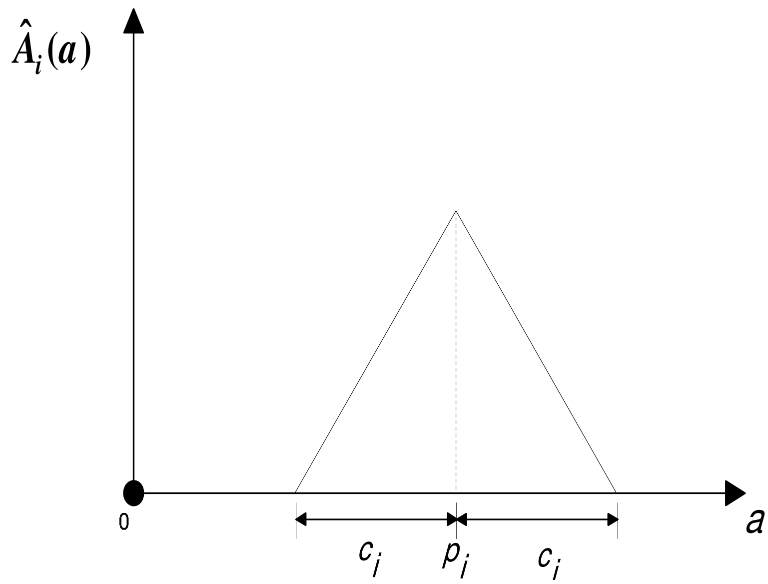

Figure 1. Membership function for the fuzzy coefficient A_i .

coefficient $\tilde{A}_i$ is shown in Figure 1.

Definition 1

A symmetric fuzzy number *A* denoted as (m_A, w_A) is defined by:

$$A(x) = \phi((x - m_A)/w_A)$$

Where reference function $\phi(x)$ satisfies: i) $\phi(x) = \phi(-x)$; ii) $\phi(0) = 1$ and iii) $\forall x_1, x_2 \in [0, \infty), x_2 > x_1 : \phi(x_1) \geq \phi(x_2)$ (Menhaj, 2006).

The membership functions for each type of A_i are assumed a triangular membership. So it can be expressed by definition 1 as:

$$\tilde{A}_i(a) = \begin{cases} 1 - \dfrac{|a - p_i|}{c_i}, & p_i - c_i \leq a \leq p_i + c_i \\ 0, & otherwise \end{cases} \tag{2}$$

Equation 1 can be written as:

$$\tilde{y} = (p_0, c_0) + (p_1, c_1)x_1 + (p_2, c_2)x_2 + \ldots + (p_n, c_n)x_n \tag{3}$$

By applying the Extension Principle (Zadeh, 1975), it implies that the membership function of fuzzy number $\tilde{y}$ is given by:

$$\tilde{y}(y) = \begin{cases} \max(\min_i \{\tilde{A}_i(a_i)\}), & \{a_i | y = f(x, a_i)\} \neq \varphi \\ 0, & otherwise \end{cases} \tag{4}$$

From Equations 3 and 4, we get:

$$\tilde{y}(y) = \begin{cases} 1 - \dfrac{\left| y - \sum_{i=1}^{n} p_i x_i \right|}{\sum_{i=1}^{n} c_i |x_i|}, & x_i \neq 0 \\ 1, & x_i = 0, y = 0 \\ 0, & x_i = 0, y \neq 0 \end{cases} \tag{5}$$

The spread of $\tilde{y}$ is $\sum_{i=1}^{n} c_i |x_i|$ and the middle of $\tilde{y}$ is $\sum_{i=1}^{n} p_i x_i$. Equation 3 can be written as:

$$\tilde{y}_j = (p_0, c_0) + (p_1, c_1)x_{1j} + (p_2, c_2)x_{2j} + \ldots + (p_n, c_n)x_{nj}$$
$$j = 1, 2, \ldots, m \tag{6}$$

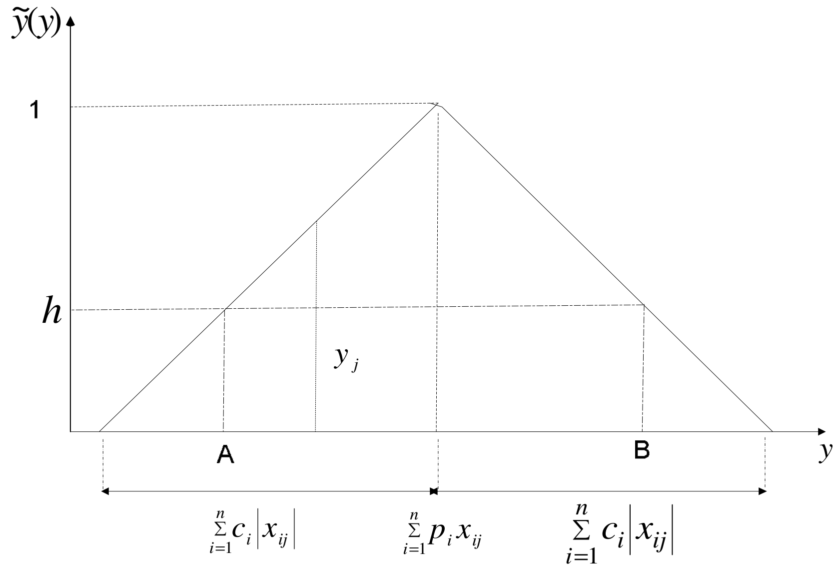

Figure 2. Fuzzy output membership function.

Where m is the number of observations.

We seek to find the coefficients $\tilde{A}_i = (p_i, c_i)$ that minimize the spread of the fuzzy output for all data sets. Equation 7 shows the objective function (Montgomery and Peck, 1982):

$$Min \sum_{j=1}^{m} \sum_{i=1}^{n} (c_0 + \sum_{i=1}^{n} c_i |x_{ij}|) \tag{7}$$

And the constraints require that each observation y_j has at least h degree of belonging to $\tilde{y}(y)$, that is (Redden and Woodall, 1996):

$$\tilde{y}(y_j) \geq h, \quad j = 1,2,...,m \tag{8}$$

The degree h is specified by the user. Figure 2 shows the membership function for the fuzzy output. Equation 8 indicates that the fuzzy output should lie between A and B as shown in Figure 2. By substituting Equation 5 into Equation 8, we obtain:

$$\begin{aligned} y_j &\geq p_0 + \sum_{i=1}^{n} p_i x_{ij} - (1-h)(c_0 + \sum_{i=1}^{n} c_i |x_{ij}|), \quad j = 1,2,...,m \\ y_j &\leq p_0 + \sum_{i=1}^{n} p_i x_{ij} + (1-h)(c_0 + \sum_{i=1}^{n} c_i |x_{ij}|), \quad j = 1,2,...,m \end{aligned} \tag{9}$$

The aforementioned analysis leads to the following linear programming problem (Tanaka et al., 1982):

$$\begin{aligned} & Min \sum_{j=1}^{m} \sum_{i=1}^{n} (c_0 + \sum_{i=1}^{n} c_i |x_{ij}|) \\ & s.t. \\ & y_j \geq p_0 + \sum_{i=1}^{n} p_i x_{ij} - (1-h)(c_0 + \sum_{i=1}^{n} c_i |x_{ij}|), \quad j = 1,2,...,m \\ & y_j \leq p_0 + \sum_{i=1}^{n} p_i x_{ij} + (1-h)(c_0 + \sum_{i=1}^{n} c_i |x_{ij}|), \quad j = 1,2,...,m \\ & c_i \geq 0, \quad p_i \geq 0 \end{aligned} \tag{10}$$

FORECASTING THE TRANSPORT ENERGY DEMAND: THE MULTI-LEVEL FLR MODEL DESIGN AND APPLICATION

Here, Iran transport energy demand is forecasted using a multi-level FLR model regarding to socio-economic and transport related indicators from 2006 to 2020. Three major variables were known to have major effects on transport energy demand: 1) The number of vehicles, 2) Population and 3) GDP. Previous studies considered explanatory variables such as gross national product (GNP), population and the total annual average vehicle kilometers, gross domestic product (GDP), urbanization rate, passenger-turnover and freight-turnover and oil price for forecasting transport energy demand (Murat and Ceylan, 2006; Zhang et al., 2009; Ceylan et al., 2008; Limanond et al., 2011; Geem, 2011). Transport energy demand has rapidly been increasing because of developments in the industrial, agricultural, transportation and commercial sectors. The population growth is the other reason for increasing of transport energy demand. The fast growth on the GDP leads to increase in the number of vehicle owners and hence to increase in energy demand in transportation sector. Therefore, in this

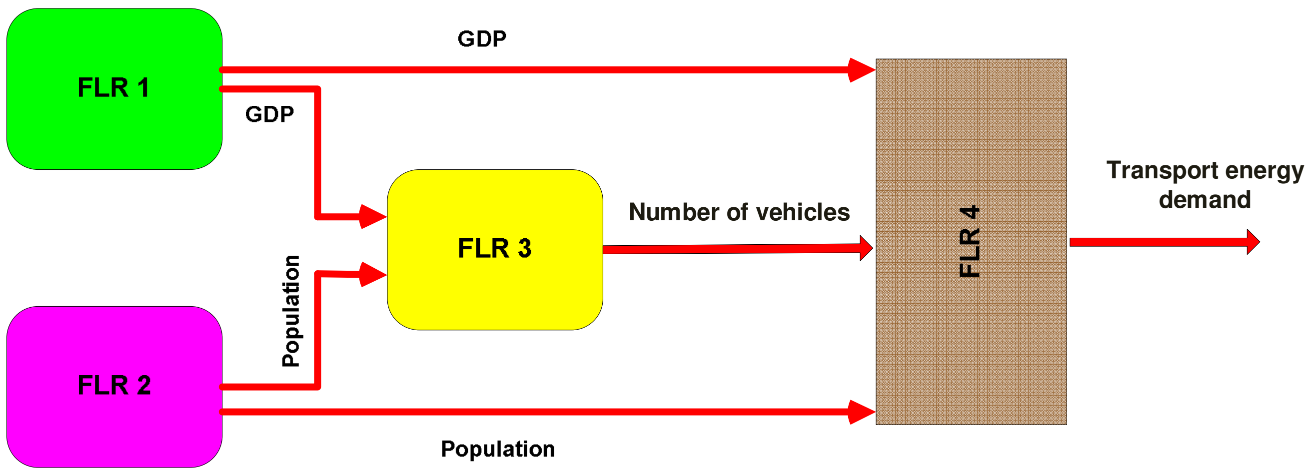

Figure 3. Structure of the designed multi-level FLR.

Table 1. FLRs inputs and output.

FLR	Inputs	Output
1	GDP in the last year GDP in the last two years	GDP
2	Population in the last year Population in the last two years	Population
3	Number of vehicles in the last year Number of vehicles in the last two years Population Ratio of GDP over population	Number of vehicles
4	Transport energy demand in the last year Transport energy demand in the last two years Number of vehicles Population Ratio of GDP over population	Transport energy demand

research, transport energy demand is analyzed based on gross domestic product (GDP), population and the number of vehicles. To estimate the energy demand model, it is needed to forecast the number of vehicles, population and GDP from 2006 to 2020. In this study, autoregressive models are used for this purpose. The structure of the designed multi-level FLR is given in Figure 3. For finding the best specification, different models have been considered. Then, the average absolute error percentage (AAEP) values of these fuzzy models are calculated and the models with minimum error are selected. The AAEP is calculated from the following equation:

$$AAEP = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{x}(i) - x(i)}{x(i)}\right| \times 100 \qquad (11)$$

Where $\hat{x}(i)$ is the estimated value and $x(i)$ is the actual value of dependent variables.

Table 1 shows the FLRs inputs and output. The main FLR (FLR 4) takes the population, the ratio of GDP over population, the number of vehicles and the transport energy demand in the last year and in the last two years as inputs and produces the transport energy demand. The population, the GDP and the number of vehicles are

Table 2. The population, GDP, the number of vehicles and transport energy demand (Iran Ministry of Energy, 2005).

Years	GDP (10^9 R)	Population	Number of vehicles	Energy demand (MBOE)
1993	258601.4	57767560	3091340	122.10
1994	259876.3	58657180	3162697	144.60
1995	267534.2	59531172	3253854	141.90
1996	283806.6	60412234	3379396	147.90
1997	291768.7	61309355	3555776	153.20
1998	300139.6	62222865	3760960	161.20
1999	304941.2	63159319	3975413	170.30
2000	320068.9	64125656	4341927	183.40
2001	330565	65126017	4741493	194.20
2002	355554	66168033	5300463	208.90
2003	379838	67256497	6084973	220.80
2004	398234.6	68399857	7027124	233.40
2005	419705	69610535	8033737	252.30

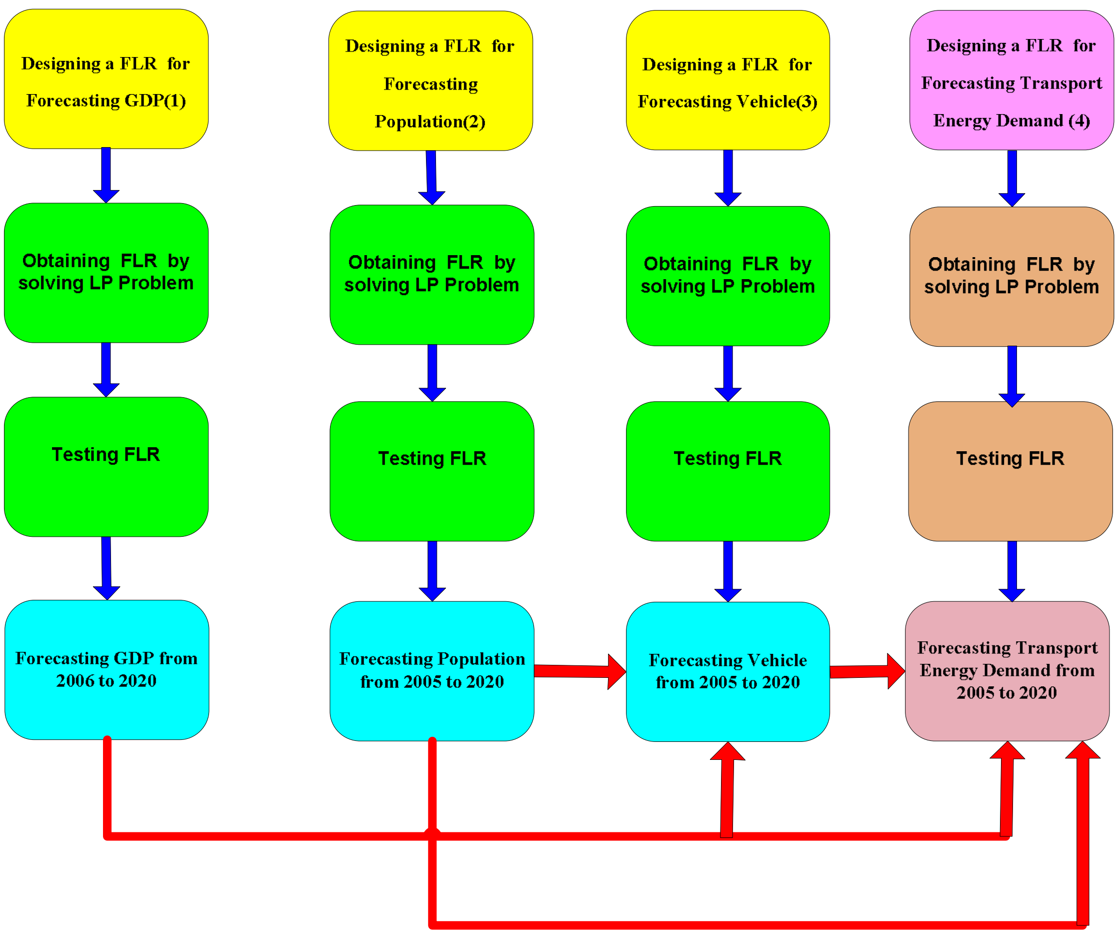

Figure 4. The framework of the forecasting.

forecasted using FLRs. Data related to transport energy modeling is collected from different sources. The GDP, population and transport energy demand are collected from Iran Ministry of Energy. The number of vehicles is collected from Iranian Fuel Conservation Organization (IFCO). Data is given in Table 2. Figure 4 shows the

Table 3. GDP, population, the number of vehicles and transport energy demand related fuzzy models.

FLR	Fuzzy Models
1	$GDP(t+1)=(p_0,c_0)+(p_1,c_1)GDP(t)+(p_2,c_2)GDP(t-1)$
2	$POP(t+1)=(p_0,c_0)+(p_1,c_1)POP(t)+(p_2,c_2)POP(t-1)$
3	$VEH(t+1)=(p_0,c_0)+(p_1,c_1)VEH(t)+(p_2,c_2)VEH(t-1)+(p_3,c_3)POP(t+1)$ $+(p_4,c_4)GDP(t+1)/POP(t+1)$
4	$EN(t+1)=(p_0,c_0)+(p_1,c_1)EN(t)+(p_2,c_2)EN(t-1)+(p_3,c_3)VEH(t+1)+$ $(p_4,c_4)POP(t+1)+(p_5,c_5)GDP(t+1)/POP(t+1)$

POP is the population, VEH is the number of vehicles, EN is the transport energy demand.

Table 4. AAEP values of GDP, population and the number of vehicles related fuzzy models.

FLR	1	2	3
AAEP	1.53%	0.00%	0.66%

Table 5. Estimated population, GDP and number of vehicles.

Years	GDP(10^9)	Population	Number of vehicles
2006	438950.99	70886844	9144579
2007	459079.53	72238383	10325503
2008	480131.08	73664691	11583765
2009	502147.97	75170966	12907068
2010	525174.47	76762556	14287711
2011	549256.87	78444995	15719644
2012	574443.59	80224039	17197377
2013	600785.28	82105699	18715589
2014	628334.89	84096270	20268982
2015	657147.81	86202360	21852236
2016	687281.98	88430922	23459994
2017	718797.98	90789278	25086867
2018	751759.18	93285151	26727437
2019	786231.85	95926691	28376270
2020	822285.30	98722508	30027925

framework of the forecasting. Fuzzy linear regression is used to build forecasting models for GDP, population, the number of vehicles and Transport Energy Demand. The fuzzy models are presented in Table 3.The study spans the time period from 1993 to 2005. This period is used to train and test the FLR models. For forecasting transport energy demand from 2006 to 2020, GDP, population and the number of vehicles are forecasted. The fuzzy linear regression is carried out to find fuzzy parameters. The average absolute error percentage (AAEP) values of GDP, population and the number of vehicles related fuzzy models can be seen in Table 4. The AAEP values are acceptable. Therefore, these fuzzy models are selected as suitable models for this study. Then, the fuzzy models are used to predict GDP, population and the number of vehicles from 2006 to 2020. The estimated socio-economic and transport related indicators are given in Table 5. The estimation of GDP, population and the

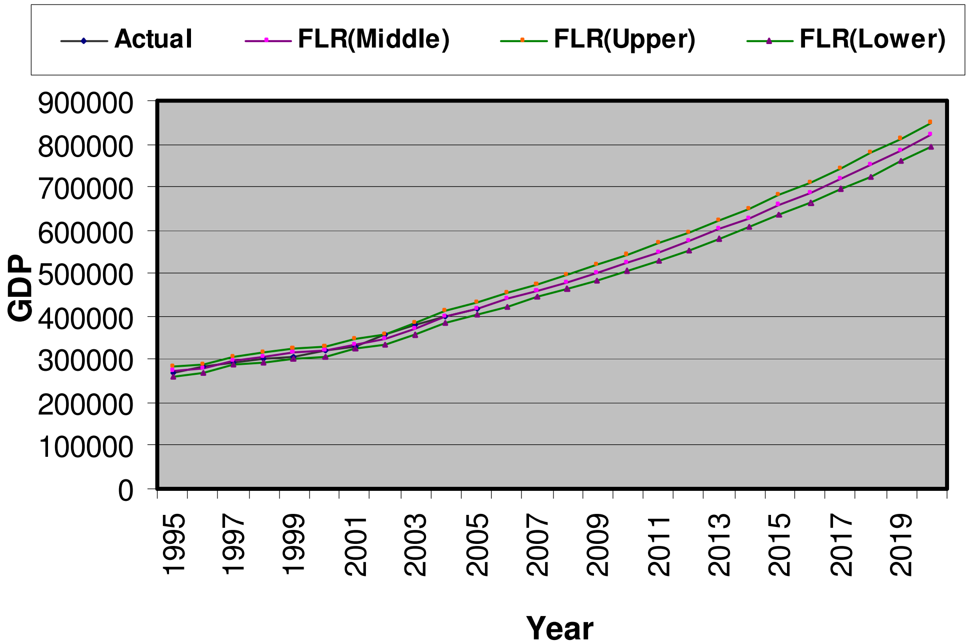

Figure 5. Estimated GDP.

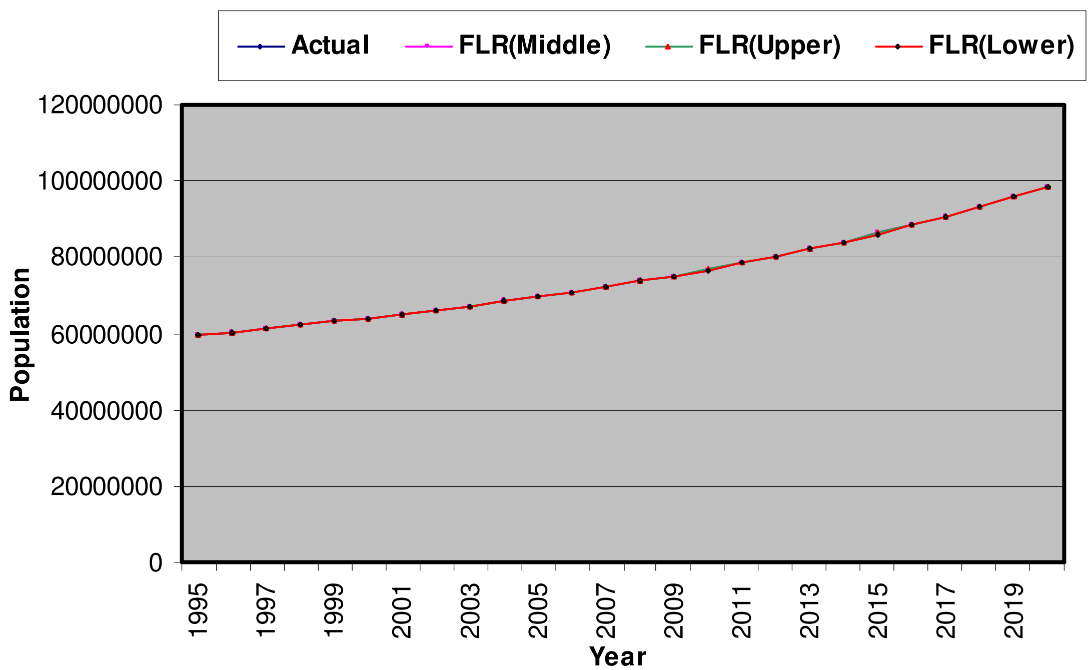

Figure 6. Estimated population.

number of vehicles are given in Figures 5, 6 and 7. These graphs show the actual data versus the FLR results. The GDP will reach to a level of 822285×10^9 *Rials*, population will be about 99 million and the number of vehicles will be almost 30 million in 2020. The fuzzy linear regression is carried out to find the transport energy demand related to fuzzy parameters. The fuzzy parameters are estimated using the socio-economic data (table 1). The AAEP value of transport energy demand related to fuzzy model is 5.72%. The AAEP value is

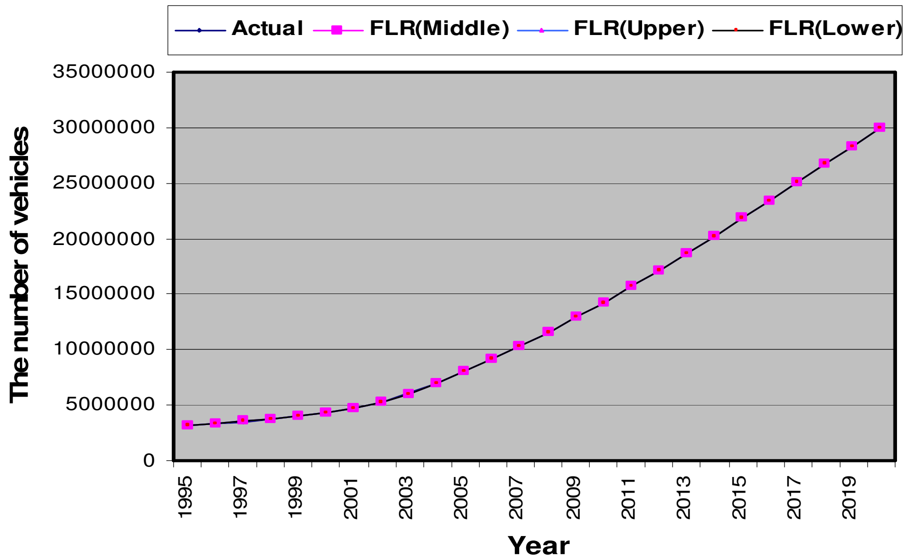

Figure 7. Estimated the number of vehicles.

Table 6. Estimated transport energy demand.

Years	Transport energy demand (MBOE)
2006	257.75
2007	269.21
2008	284.96
2009	302.57
2010	321.94
2011	342.79
2012	365.03
2013	388.59
2014	413.47
2015	439.70
2016	467.29
2017	496.32
2018	526.82
2019	558.87
2020	592.55

acceptable. Therefore, this fuzzy model selected as suitable model for forecasting the transport energy demand. Then, the estimated fuzzy parameters are used to predict the transport energy demand till the year 2020. The results can be seen in Table 6.

The estimated values are given in Figure 8. Transport energy demand will reach to a level of 592 MBOE in 2020.

Conclusion

This study discussed the transport energy demand for the period of 1993 to 2005 based on GDP, population and the number of vehicles as independent variables. The transport energy demand was forecasted for the time span 2006 to 2020. The concept of fuzzy linear regression was described in detail. Then fuzzy models were built for the transport energy demand, GDP, population and the number of vehicles in Iran. After that, fuzzy models related fuzzy parameters were obtained. The following conclusions were obtained from this study: 1) Transport energy demand will reach to a level of 592 MBOE in 2020; 2) GDP will reach to a level of 822285×10^{9} *Rials* in 2020; 3) Population will be about 99

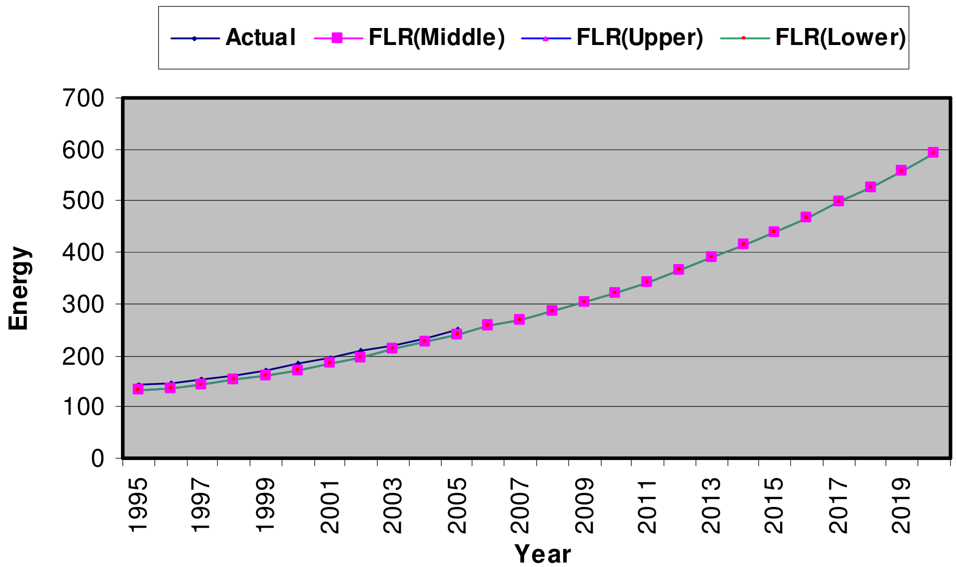

Figure 8. Estimated transport energy demand.

million in 2020; 4) The number of vehicles will be almost 30 million in 2020. Future studies can take into account the other various variables to estimate the transport energy demand. The benefits of the energy savings when the control measures are imposed to restrict private car usage are currently under investigation.

REFERENCES

Akkuş H, Asilturk I (2011). Predicting surface roughness of AISI 4140 steel in hard turning process through artificial neural network, Fuzzy Logic and Regression Models. Sci. Res. Essays., 6(13): 2729-2736.

Al-Kandari AM, Soliman SA, El-Hawary ME (2004). Fuzzy short-term electric load forecasting. Electr. Pow. Energ. Syst. 26(2): 111–122.

Bel PM, Wang H (1977). Fuzzy linear regression models for assessing risks of cumulative trauma disorders. Fuzzy Set. Syst., 92(3): 17-340.

Ceylan HU, Ceylan HA, Haldenbilen S, Baskan O (2008). Transport energy modeling with meta-heuristic harmony search algorithm, an application to Turkey. Energ. Policy, 36: 2527-2535.

Chang PT (1997). Fuzzy seasonality forecasting. Fuzzy Set. Syst., 90(1): 1-10.

Geem ZW (2011). Transport energy demand modeling of South Korea using artificial neural network. Energ. Policy, 39: 4644-4650.

Heshmaty B, Kandel A (1985). A Fuzzy linear regression and its applications to forecasting in uncertain environment. Fuzzy Set. Syst., 15(2): 159-191.

Limanond T, Jomnonkwao S, Srikaew A (2011). Projection of future transport energy demand of Thailand. Energ. Policy, 39: 2754-2763.

Menhaj MB (2006). Fuzzy Computations. Amirkabir University of Technology, Iran.

Montgomery DC, Peck EA (1982). Introduction to Linear Regression Analysis. Wiley, New York.

Murat YS, Ceylan H (2006). Use of artificial neural networks for transport energy demand modeling. Energ. Policy, 34 3165-3172.

Redden DT, Woodall WH (1996). Further examination of fuzzy linear regression. Fuzzy Set. Syst., 79(2): 203-211.

Shakouri HG, Nadimi R, Ghaderi SF (2008). Hybrid TSK-FR model to study short-term variations of the electricity demand versus the temperature changes. Expert Syst. Appl., 36(2) : 1765-1772.

Soliman SA, Alammari RA, El-Hawary ME (2003). Frequency and harmonics evaluation in power networks using fuzzy regression technique. Electr. Pow. Syst. Res., 66(2): 171-177.

Srivastava AK, Nema AK (2008). Forecasting of solid waste composition using fuzzy regression approach: a case of Delhi. Int. J. Environ. Waste Manage., 2(1/2): 65 - 74.

Tanaka H, Uejima S, Asai K (1982). Fuzzy linear regression model. IEEE Tans System Man Cybernetics 12: 903- 907.

Tanaka H, Uejima S, Asai K (1982b). Fuzzy linear regression model. IEEE T. Syst. Man Cyb. 10: 2933–2938.

Yalpir S, Özkan G (2011), Fuzzy logic methodology and multiple regressions for residential real-estates valuation in urban areas. Sci. Res. Essays., 6(12): 2431-2436.

Zadeh LA (1965). Fuzzy sets. Inf. Control., 8(3): 338-353.

Zadeh LA (1975). The concept of a linguistic variable and its application to approximate reasoning-I. Inf. Sci., 8: 199-249.

Zhang M, Mu H, Li G, Ning Y (2009). Forecasting the transport energy demand based on PLSR method in China. Energy, 34: 1396-1400.