University of Economics in Katowice

Journal of Economics and Management

ISSN 1732-1948 Vol. 36 (2) • 2019

Mohsen Tavan

Department of Economics
Faculty of Literature and Humanities
Persian Gulf University, Bushehr, Iran
mohsentavan20@gmail.com

Estimation and forecast of carbon dioxide emission in Iran: Introducing a new hybrid modelling

doi: 10.22367/jem.2019.36.08

Accepted by Editor Ewa Ziemba | Received: April 23, 2018 | Revised: August 13, 2018; December 11, 2018; January 24, 2019; February 2, 2019 | Accepted: February 7, 2019.

Abstract

Aim/purpose – The purpose of this paper is to introduce a new hybrid modelling to predict carbon dioxide emissions in order to make the correct decision to reduce air pollution in Iran. While there are not many data available for some variables, in this modelling, the goal is to make accurate predictions even with low data.
Design/methodology/approach – In the present paper, CO2 emissions in Iran in the period of 1980-2014 was predicted using three models of Auto-Regressive Distributed Lag (ARDL), Fuzzy Linear Regression (FLR) and hybrid model based on combination of ARDL and FLR models, and then the prediction accuracy of the models is compared.
Findings – Comparing prediction accuracy of models showed that the Fuzzy Auto-Regressive Distributed Lag (FARDL) is more accurate than the initial patterns for predicting carbon dioxide emissions. Finally, the results showed that GDP and energy consumption has a positive and significant correlation with carbon dioxide emissions in short run. Also, the carbon dioxide emission indicated a low elasticity towards GDP and low energy consumption.
Research implications/limitations – When the number of data is low, the FARDL model provides a more accurate prediction than ARDL and FLR Models. FARDL's combined model reduces the problems that exist in the ARDL and FLR models. One of the problems with the ARDL model is the need for many tests; the problem with the fuzzy regression model is also the high fuzzy distance length that makes decision making difficult. The FARDL model eliminates these constraints as much as possible.
Originality/value/contribution – This paper has been able to confirm that Fuzzy Auto-Regressive Distributed Lag (FARDL) is more accurate than the initial patterns for predicting carbon dioxide emissions.

Keywords: forecast, carbon dioxide, Fuzzy Auto-Regressive Distributed Lag, Iran.
JEL Classification: Q53, Q41, P48, P18.

1. Introduction

The exact prediction of the future is one of the most important steps that will lead to proper management and planning and, hence, to the reduction of future losses. If the prediction is not accurate, it can cause irreparable damage, so creating modelling for an accurate forecast is essential. Considering that there is not much data available in some studies, this study attempts to create a model that can accurately predict even with low data.

As in the study of Tseng, Tzeng, Yu, & Yuan (2001), the ARIMA (Auto--Regressive Integrated Moving Average) and FLR (Fuzzy Linear Regression) models have been used to create the FARIMA (Fuzzy Auto-Regressive Integrated Moving Average), model. In this study, two ARDL (Auto-Regressive Distributed Lag) and FLR models have been used to create the FARDL (Fuzzy Auto-Regressive Distributed Lag) model. As in the research of Tseng et al. (2001), here were also compared the accuracy of prediction of a new hybrid model with two basic models. Due to the importance of air pollution, the carbon dioxide emission variable is used in this study to evaluate the accuracy of models prediction.

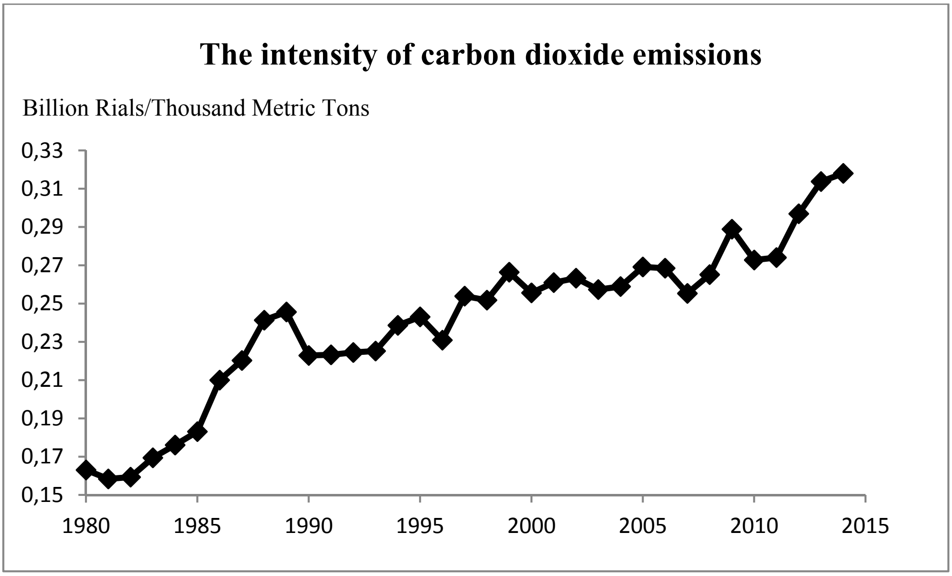
Carbon dioxide is regarded as one of the most important greenhouse gases that causes weather changes and global warming. Intergovernmental Panel on Climate Change (IPCC) linked the temperature anomaly with the cumulative total of greenhouse gas (GHG) emissions since the Industrial Revolution, calculating carbon budgets. The quantity of carbon dioxide in the atmosphere is determined by the amounts supplied and withdrawn from three other great reservoirs: oceans, rocks, and living organisms. The Paris Agreement of December 2015 committed to acting towards the objectives of keeping a global temperature rise this century well below 2 °C above pre-industrial levels, and to pursue efforts to further limit the temperature increase to 1.5 °C. A very important anthropogenic source of $CO2$ emissions is energy production from sources such as coal and oil combustion. There is a relationship between the wealth of society, the state budget and the amount of energy produced. This is obviously related to environmental pollution during the production of energy from various natural sources (coal, gas, oil) on a massive scale.

Nowadays, along with the increase in GDP of the country, attention to environmental protection and reduction of carbon dioxide emission is one of the important goals of policymakers. Failure to pay attention to the environment will increase pollution and environmental degradation in accordance with the growth in the country's GDP. Among the factors that can reduce carbon dioxide one can name reducing environmental pollution, applying strict regulations to protect the environment, production and import of machinery and equipment that meet the required standards and save the energy. Cheap energy in oil-rich countries, including Iran, and low energy consumption efficiency have resulted in excessive consumption of energy and increased environmental pollution. Carbon dioxide emission causes air pollution, increases global warming, and leads to environmental degradation. In addition, carbon dioxide emission is a measure of pollution and environmental degradation. Carbon dioxide is one of the most important greenhouse gases that causes global warming. Industrialisation and intensive use of fossil fuels, such as oil and gas, release significant amounts of carbon dioxide gas into the atmosphere.

Over the past years, many studies have been carried out on environmental pollution considering various factors such as economic growth, energy consumption, and population, as independent variables for the emission of carbon dioxide.

The intensity of carbon dioxide emissions in Iran (carbon dioxide emission to GDP) for the years 1980-2014 is shown in Figure 1.

Figure 1. Intensity of carbon dioxide emissions in Iran during the study period (1980-2014)



Source: Adapted from: Energy Information Administration [EIA] (2016); Ministry of Energy Iran (2010).

As presented in Figure 1, the intensity of carbon dioxide emissions during the study period (1980-2014) was constantly increasing with limited fluctuations. The increase in carbon dioxide emission during the study shows that it rose faster than the country's gross domestic product (GDP), probably because of excessive consumption of energy due to cheap energy in Iran as well as low-performance machinery and low-quality equipment in different parts of the country.

In this study, Microfit and WinQSB software was used for modelling and forecasting. In this respect, it was attempted to create a new hybrid model with more accuracy, using the ARDL and FLR models. Section 2 presents the literature review and the methodology is presented in section 3. Section 4 shows the application of the models and results. Section 5 discusses the model results, limitations and conclusions and finally provides recommendations.

2. Literature review

Combined models are used to increase the accuracy of prediction and reduce the problems and limitations of the base models. For example, in the study of Tseng et al. (2001), the FARIMA combination model is used to increase the accuracy of prediction and reduce the problems and limitations of the two ARIMA and FLR models. In this research, carbon dioxide emissions are estimated using the ARDL model, then in the FLR model; the estimation coefficients are used to reduce the fuzzy distance and increase prediction accuracy, and the new FARDL model is created.

In most previous studies, GDP and energy consumption variables have been used to estimate carbon dioxide emissions in Iran, and only these variables are used for estimation and prediction. Since a new model was used in this study, only the internal and foreign studies related to carbon dioxide emissions are presented in the following sections.

2.1. Foreign studies

Soytas, Sari, & Ewing (2007) studied the relationship between energy consumption, income, and carbon emissions in the United States and concluded that there is a positive and significant relationship between carbon emissions and energy consumption, while there is no meaningful relationship between income and carbon emissions. They also reported that income growth in the United States cannot be the cause of the environmental problems in the country.

Soytas & Sari (2009) investigated the causal relationship between three variables of energy consumption, economic growth, and carbon emissions in Turkey. They used the variables of human resources, capital, and economic growth, and carbon emissions to examine the relationship between economic growth and carbon emissions. They extracted a unilateral relationship between carbon emissions and energy consumption, while such link between carbon emissions and income has not been confirmed. Finally, they acknowledged that a reduction in carbon emissions would not reduce economic growth in Turkey.

Acaravci & Ozturk (2010) examined the relationship between energy consumption, carbon dioxide emissions, and economic growth using the co-integration approach and the ARDL model in 19 European countries based on the Kuznets environment model. Their results show a long-run relationship among the three studied variables only for 7 countries.

Hamit-Haggar (2012) reported that there is a positive and significant relationship between energy consumption in the Canadian industrial sector and greenhouse gas emissions and by increasing energy consumption in the industrial sector, greenhouse gas emissions will increase. In addition, he proved the Kuznets curve hypothesis.

2.2. Internal studies

Mohammad Bagheri (2010) explored the relationship between GDP, energy consumption, and carbon dioxide emissions. He concluded that carbon dioxide emission has a low elasticity towards the GDP. In addition, carbon dioxide elasticity towards energy consumption is similar to and about 1 in the short and long terms.

Sadeghi, & Islami Ordoghly (2011) studied the economic growth and environmental pollution in the Kyoto Protocol countries. They showed a significant and positive correlation between energy consumption and carbon dioxide emissions, and a nonlinear relationship between carbon dioxide emissions and GDP per capita. In addition, based on the results, there is a one-way Granger causality relationship from the use of fossil fuels and carbon dioxide emissions to economic growth in the long run and a one-way short-run causality relationship from carbon emissions to fossil energy consumption.

Moghaddasi & Golriz Ziaee (2011) investigated the causal relationship between carbon dioxide emissions and GDP based on mixed data for five groups of countries with different per capita income during the period of 1960-2007. They concluded that there is a causal relationship between these two variables using the co-integration test and error correction models. Additionally, they showed that in all studied countries, with increased income, carbon dioxide emissions are also increased.

Lotfalipour, Fallahi, & Ashena (2011) in the article entitled 'Investigating the relationship between carbon dioxide emissions and economic growth, energy and trade in Iran' used the data related to years 1966-2007 and causality test based on error correction model. The results showed that causation is associated with carbon growth, economic growth, fossil energy consumption, and commercial freedom, but the inverse relationship is not confirmed.

Sadeghi & Ebrahimi (2013) in a study entitled 'Impact of financial development, GDP, and energy consumption on environmental pollution' used the data of the years 1971-2008 and the ARDL model. They concluded that financial development, GDP, energy use, and trade liberalisation have a positive and significant impact on carbon dioxide emissions.

Jafari Samimi & Mohammadi Khayareh (2014) in their study entitled 'Short--term and long-term relationships of carbon dioxide emissions, energy use, and economic growth' used data from 1979-2010 and applied the co-integration Bound Test approach. The results indicate the existence of a long-term relationship between variables and the existence of a one-way causality relationship from the per capita GDP to per capita energy consumption and per capita carbon emissions.

3. Research methodology

3.1. Auto-Regressive Distributed Lag

Today, the use of Auto-Regressive distributed (ARDL) model has been widespread. This model was originally described by Pesaran & Shin (1998) and Pesaran et al. (2001). In the Auto-Regressive model with extended lags for each of the variables, optimal lags are selected using the Schwarz–Bayesian criteria (SBC, Akaic (AIC), Hanakin Quinn (HQC), or the correction coefficient (Pahlavan & Dahmardeh, 2007). To reduce the bias related to the estimation of pattern coefficients in small samples, it is best to use a model that does not take into account the number of lags for variables; however, in samples less than 100, the Schwartz–Bayesian criterion prevents losing a degree of freedom. This criterion allows for the determination of lags and, as a result, provides an estimate of a greater degree of freedom (Pesaran & Shin, 1996). The Schwarz–Bayesian criterion is also used to select optimal lags.

In the Engle–Granger method, the obtained estimates in small samples are biased due to the lack of consideration of the short-term dynamic responses between the variables. The Engle–Granger method is based on the presumption of the existence of a co-integration vector. Under conditions where more than one Co-Integrate vector exists, the use of this method will result in inefficiency (Pesaran & Smith, 1998). To overcome these defects, Johansen & Juselius (1992) suggested a Maximum Likelihood Ratio method for co-integration test and extracting co-integration vectors. The Johansen–Juselius co-integration method cannot be useful because all model variables have the same degree of reliability (Zaranezhad & Saadatmehr, 2007).

ARDL, in addition to calculating long-term relationships, also allows for simultaneous calculation of dynamic and short-term relationships. The short--term moderation rate in each period can also be calculated to achieve long-term equilibrium. Unlike the Johansson–Juselius method, where all variables must be of the same order of magnitude, it is not necessary to determine the degree of stability of the variables. By choosing appropriate lags for the variables, a suitable model can be selected. Another important advantage of this method among the co-integration methods is that this method can be applied regardless of the argument that the variables of the model are either stable or unstable; in other words, in this method, it is not necessary to divide variables into the variables full of 0 and 1 degrees. Besides, this pattern gains unbiased estimators in small samples and thus avoids problems such as autocorrelation, bias, and inefficiency. Most importantly, because some variables are stable and some non-stationary, the ARDL method is suitable to investigate the relationship between level variables (Abbasinejad & Goodarzi Farahani, 2013).

The mathematical structure of the ARDL (p, qi) model is as follows:

$$\alpha(L,p)y_t = \sum_{i=1}^{m} \beta_i(L,q_i)x_{it} + c'w_t + \varepsilon_t$$

$$\alpha(L,p) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$$

$$\beta_i(L,q_i) = \beta_{i0} + \beta_{i1}L + \beta_{i2}L^2 + \ldots + \beta_{iq}L^q \tag{1}$$

$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \beta_{i0}x_{it} + \beta_{i1}x_{i(t-1)} + \ldots + \beta_{iq}x_{i(t-q)} + c'w_t + \varepsilon_t \tag{2}$$

t = 1,2,…,n i = 1,2,…,m

where:
y_t is the dependent variable,
x_t is the independent variable,
L is the l operator ($y_{t-1} = Ly_t$), α is the dependent variable parameters,
β is the parameters of the independent variables,
p is the number of lags of the dependent variable,
q is the number of lags of the independent variables,
w_t is the vector S × 1, which represents the predetermined variables in model including the width of the source, the virtual variables, the time trend, and other exogenous variables, ε_t is pure error term and random variable assumed with a normal distribution with zero mean, and σ^2 variance (Noferesti, 1999). In the ARDL method, a two-step approach can be used to examine long-term relationships. In the first stage, there is a long-term relationship between the examined variables (Pesaran et al., 2001). At this point, there are two ways to check that the long-term relationship is not false:

In the first method, after estimating the dynamic model of ARDL, the following hypothesis is tested:

$$H_0 = \sum_{l=1}^{p} \beta_l - 1 \geq 0$$
$$H_1 = \sum_{l=1}^{p} \beta_l - 1 \leq 0 \tag{3}$$

The zero hypothesis indicates that there is no co-integration or long-term relationship. To perform the test, which is presented by Banerjee, Dolado, & Mestre (1993), the number of coefficients with the dependent variable lag must be calculated and divided by the sum of the standard deviations of these coefficients. The t-test statistic is as follows:

$$t = \frac{\sum_{t=1}^{p} \hat{\beta}_i - 1}{\sum_{l=1}^{p} \delta_{\hat{\beta}_l}} \tag{4}$$

If the absolute value of the t-statistic obtained from the absolute magnitudes of the critical values provided by Banerjee et al. (1993) are higher at 95% confidence level and the zero hypothesis, based on the absence of co-integration and the existence of a long-term relationship, is accepted.

3.2. Fuzzy Linear Regression Model

The theory of the fuzzy set was firstly established by Lotfi A. Zadeh (1965) and fifteen years later fuzzy regression was discussed by Fukami, Mizumoto, & Tanaka (1980). Since the concept of the error term is used in the regression and the integrated moving autoregressive models, in the modelling of these models, all assumptions for the error term should be taken into account. Tanaka et al. (1984) presented fuzzy regression, which is an interval prediction model to avoid modelling error. A generalised fuzzy regression is a regular regression that is used to compute the functional relationship between independent variables and dependent variables in a fuzzy environment and does not include the error term (Zimmermann, 1996). In this model, there are two upper and lower bounds where the actual values of the dependent variable are between these two limits; therefore, the fuzzy linear regression is written as follows:

$$\overset{\approx}{Y} = \overset{\approx}{A}_0 + \overset{\approx}{A}_1 x_1 + \ldots + \overset{\approx}{A}_m x_m \tag{5}$$

Where $\overset{\approx}{Y}$ is fuzzy dependent variable, x_i is independent variables, m is the number of independent variables, and $\overset{\approx}{A}_i$ is the fuzzy coefficient of the ith independent variable. $\overset{\approx}{A}_i$, which presents the form of the fuzzy numbers of Dubois & Prade (1980), is defined as follows:

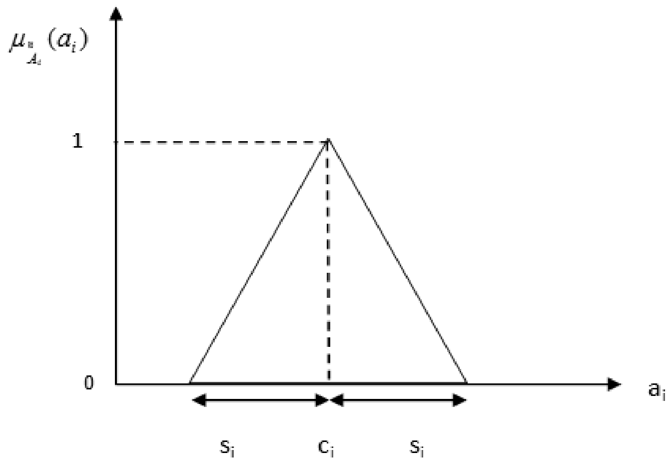
$$\mu_{\overset{\approx}{A}_i}(a_i) = L\left\{\left(\frac{|c_i - a_i|}{s_i}\right)\right\} \tag{6}$$

$\mu_{\overset{\approx}{A}_i}(a_i)$ is the membership function of a fuzzy set and expresses the factors of $\overset{\approx}{A}_i$. Also, c_i and s_i represent the centre and width of the membership function, respectively. As a result, the membership function of regression model coefficients, which is in the form of symmetric fuzzy triangular numbers, is defined as follows:

$$\mu_{\overset{\approx}{A}_i}(a_i) = \begin{cases} 1 - \dfrac{|c_i - a_i|}{s_i} & c_i - s_i \le a_i \le c_i + s_i \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$\mu_{\tilde{\tilde{A}}_i}(a_i)$ is the membership function of a fuzzy set and expresses the factors of $\tilde{\tilde{A}}_i$. Also, c_i and s_i represent the centre and width of the membership function, respectively (Tseng et al., 2001). The objective of the fuzzy regression model is to determine the optimal values of fuzzy coefficients, so that the fuzzy output membership degree is larger for all points than a given value of h. The choice of h is effective on the expansion of fuzzy factors and is determined by the user. Increasing (1-h) increases the fuzzy values (Yen, Ghoshray, & Roig, 1999). The form of the membership function of fuzzy coefficients is also plotted by Zimmermann (1996), which is as follows:

Figure 2. The membership function of fuzzy coefficients



As shown in Figure 2, the membership value of the fuzzy set can be a real arbitrary value between zero and one, and the closer the value is to one, the degree of membership of the element of a_i is greater than the fuzzy set, and if its value is zero, then the element a_i does not belong to the fuzzy set at all.

Relation (5) can be written as follows:

$$\tilde{\tilde{Y}} = (c_0, s_0) + (c_1, s_1)x_1 + \ldots + (c_n, s_n)x_m \tag{8}$$

The extension principle is one of the fundamental concepts of the theory of fuzzy sets that was expressed by Zadeh (1965) to generalise definite mathematical concepts to fuzzy concepts. Given the extension principle around the centre, the fuzzy output variable membership function is as follows:

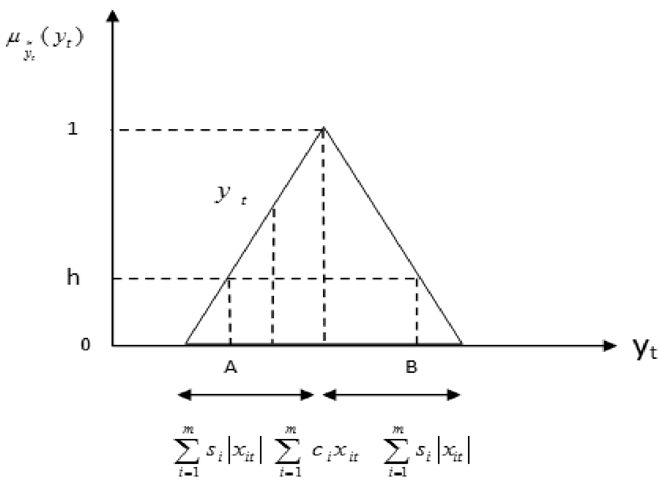
$$\mu_{\tilde{\tilde{y}}_t}(y_t) = \begin{cases} 1 - \dfrac{\left| y_t - \sum_{i=1}^{m} c_i x_{it} \right|}{\sum_{i=1}^{m} s_i \left| x_{it} \right|} & x_t \neq 0 \\ 1 & x_t = 0 \quad y_t = 0 \\ 0 & x_t = 0 \quad y_t = 0 \end{cases} \tag{9}$$

As already mentioned, the value of the membership function of the fuzzy output variable must be greater than the specified value of h and h is determined by the researcher.

$$\mu_{\tilde{\tilde{y}}_t}(y_t) \geq h \tag{10}$$

The following figure shows that the fuzzy output membership function must be between two values of A and B.

Figure 3. Membership function of the fuzzy output



If (9) is replaced in (10):

$$1 - \frac{\left| y_t - \sum_{i=1}^{m} c_i x_{it} \right|}{\sum_{i=1}^{m} s_i \left| x_{it} \right|} \geq h \tag{11}$$

After simplification, the following limitations are achieved:

$$c_0 + \sum_{i=1}^{m} c_i x_{it} - (1-h)\left[s_0 + \sum_{i=1}^{m} s_i x_{it} \right] \le y_t$$
$$c_0 + \sum_{i=1}^{m} c_i x_{it} + (1-h)\left[s_0 + \sum_{i=1}^{m} s_i x_{it} \right] \ge y_t \quad (12)$$

t = 1,2,…n

and thus the problem of finding fuzzy regression parameters through linear programming is as follows (Tanaka & Ishibuchi, 1992):

$$\min : \sum_{t=1}^{n} s_0 + \sum_{i=1}^{m} \sum_{t=1}^{n} s_i x_{it}$$
$$s.t.$$
$$c_0 + \sum_{i=1}^{m} c_i x_{it} - (1-h)\left[s_0 + \sum_{i=1}^{m} s_i x_{it} \right] \le y_t \quad (13)$$
$$c_0 + \sum_{i=1}^{m} c_i x_{it} + (1-h)\left[s_0 + \sum_{i=1}^{m} s_i x_{it} \right] \ge y_t$$

$$s_i \ge 0 \, , c_i \ge 0 \, , \; 0 \le h \le 1 \, , \; i = 1, 2, \ldots, m \; , \; t = 1, 2, \ldots, n$$

As can be seen, in high linear programming, the width or fuzzy amount is minimised and the constraints also indicate what the dependent variable should be. After obtaining fuzzy coefficients, the approach proposed by Ishibuchi & Tanaka (1988) is used to reduce the range of prediction. In this method, constraints related to the values of the dependent variable that are placed on the upper and lower limits are deleted and new coefficients are re-obtained. Finally, the coefficients are obtained in the regression equations, and the upper and lower bounds of the dependent variable are obtained in a fuzzy manner.

3.3. Fuzzy Auto-Regressive Distributed Lag

In this model, following a Fuzzy Auto-Regressive Integrated Moving Average model (FARIMA) (Tseng et al., 2001). The components of the ARDL model are used in the fuzzy linear regression model. In this model, the coefficients and errors terms of the ARDL are used in the fuzzy linear regression model. The self-explanatory pattern with broad lags is as follows:

$$\alpha(L, p) y_t = \sum \beta_i (L, q_i) x_{it} + c' w_t + \varepsilon_t \quad (14)$$

$$\alpha(L,p)=1-\alpha_1 L-\alpha_2 L^2-\ldots-\alpha_p L^p$$

$$\beta_i(L,q_i)=\beta_{i0}+\beta_{i1}L+\beta_{i2}L^2+\ldots+\beta_{iq}L^q$$

$$y_t=\alpha_1 y_{t-1}+\ldots+\alpha_p y_{t-p}+\beta_{i0}x_{it}+\beta_{i1}x_{i(t-1)}+\ldots+\beta_{iq}x_{i(t-q)}+c'w_t+\varepsilon_t \quad (15)$$

The above model in the fuzzy form is as follows:

$$\overset{\approx}{y}_t=\overset{\approx}{A}_1\, y_{t-1}+\ldots+\overset{\approx}{A}_p\, y_{t-p}+\overset{\approx}{B}_{i0}\, x_{it}+\overset{\approx}{B}_{i1}\, x_{i(t-1)}+\ldots+\overset{\approx}{B}_{iq}\, x_{i(t-q)}+\overset{\approx}{A}_0+\varepsilon_t \quad (16)$$

According to the extension principle around the centre, the fuzzy output variable membership function is as follows:

$$\overset{\approx}{y}(y)=\begin{cases}1-\dfrac{\left|y-c_0-\sum_{i=1}^{m}c_i x_{ij}\right|}{s_0+\sum_{i=1}^{m}s_i\left|x_{ij}\right|} & x_i\neq 0\\ 1 & x_i=0, y=0\\ 0 & x_i=0, y\neq 0\end{cases} \quad (17)$$

The membership function of the fuzzy output variable must be greater than the specified value of h.

$$\overset{\approx}{y}(y)\geq h \quad (18)$$

The above relation is also the relation (10).

Relation (17) is replaced in (18):

$$1-\frac{\left|y-c_0-\sum_{i=1}^{m}c_i x_{ij}\right|}{s_0+\sum_{i=1}^{m}s_i\left|x_{ij}\right|}\geq h \quad (19)$$

After simplification, the model is as follows:

$$\min : \sum_{t=1}^{n} s_0 + \sum_{i=1}^{p}\sum_{t=1}^{n} s_i \left|\alpha_i\right|\left|y_{t-i}\right| + \sum_{j=1}^{q}\sum_{t=1}^{n} s_j \left|\beta_j\right|\left|x_{jt}\right|$$

$s.t.$

$$c_0 + \sum_{i=1}^{p}\sum_{t=1}^{n} c_i y_{t-i} + \sum_{j=1}^{q}\sum_{t=1}^{n} c_j x_{jt} + \varepsilon_t - (1-h)\left[s_0 + \sum_{i=1}^{p}\sum_{t=1}^{n} s_i y_{t-i} + \sum_{j=1}^{q}\sum_{t=1}^{n} s_j x_{jt} \right] \le y_t \quad (20)$$

$$c_0 + \sum_{i=1}^{p}\sum_{t=1}^{n} c_i y_{t-i} + \sum_{j=1}^{q}\sum_{t=1}^{n} c_j x_{jt} + \varepsilon_t + (1-h)\left[s_0 + \sum_{i=1}^{p}\sum_{t=1}^{n} s_i y_{t-i} + \sum_{j=1}^{q}\sum_{t=1}^{n} s_j x_{jt} \right] \ge y_t$$

$t = 1, 2, 3 \ldots, n$

The above relationships are as in (13).

Finally, after obtaining the fuzzy coefficients, the model is as follows:

$$\begin{aligned} \overset{\approx}{y}_t &= (c_1, s_1) y_{t-1} + \ldots + (c_p, s_p) y_{t-p} + (c_{b0}, s_{b0}) x_{it} + (c_{b1}, s_{b1}) x_{i(t-1)} \\ &+ \ldots + (c_{bq}, s_{bq}) x_{i(t-q)} + (c_0, s_0) + \varepsilon_t \end{aligned} \quad (21)$$

The relation above is also the relation (8).

$\overset{\approx}{y}_t$ is the fuzzy dependent variable, x_i is the ith explanatory variable, p is the number of lags of the dependent variable, q is the number of lags of the independent variable, t is the time, $(c,s)_t$ is the fuzzy coefficient of the ith variable, and ε_t is the error term.

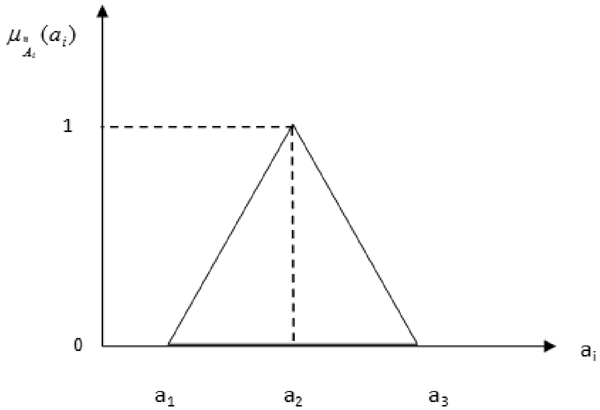
3.4. Defuzzification

The coefficients of the fuzzy regression model are fuzzy numbers. So, the output variable is also fuzzy. Consequently, non-fuzzy numbers must be converted to estimate the accuracy of prediction. There are several methods for defuzzification. In the method of mean value, the left and right separations in addition to the simplicity of all membership function are used for defuzzification (Amiri, 2010). In this study, the mean value method, which is expressed by the following equation, is used:

$$\overset{\approx}{A} = (a_1, a_2, a_3) \quad (22)$$

where $\overset{\approx}{A}$ is a triangular fuzzy number and a1, a2, and a3 are the left-hand, centre, and right-hand sides of the triangle, respectively.

Figure 4. Triangular fuzzy number



$$S(\tilde{\tilde{A}}) = \frac{1}{2}(S_L(\tilde{\tilde{A}}) + S_R(\tilde{\tilde{A}})) = \frac{1}{2}(a_2 - \int_{a_1}^{a_2} f_A(x) + a_2 + \int_{a_2}^{a_3} f_A(x)) = \frac{a_1 + 2a_2 + a_3}{4} \tag{23}$$

$S_L(\tilde{\tilde{A}})+S_R(\tilde{\tilde{A}})$ are the left and right regions of triangular fuzzy number, respectively (Amiri, 2010).

3.5. Forecast accuracy of evaluation criteria

In forecast methods, data are divided into two parts. The first part is used to estimate the model and the second part is used for comparison. Many methods such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are used to measure the prediction accuracy of the model (Zaranezhad, Ebrahimi, Raoofi, & Kiani, 2012).

RMSE represents the difference between the predicted value of the model or the statistical estimator and the actual value. This criterion is a good tool for comparing forecast errors with a dataset, while it is not applicable for comparing multiple datasets. This difference is called residual, which is computed using RMSE.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n}} \tag{24}$$

Another common criterion for evaluating the forecasts is the MAPE. The advantage of using the MAPE to forecast errors is that it can be used to compare the forecast of series with different scales because this index is not scale--dependent. The MAPE is presented as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(Y_i - \hat{Y}_i)}{Y_i}\right| * 100 \tag{25}$$

The MAE is presented as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|(Y_i - \hat{Y}_i)\right| \tag{26}$$

The higher the values are, the less the prediction error.

Using R^2, the explanatory power of the regression variables can be analysed. The more the value, the more the explanatory power of the variables would be.

$$R^2 = \frac{\left[\sum_{i=1}^{n}(\hat{Y}_i - \check{Y}_i)(Y_i - \mathring{Y}_i)\right]^2}{\sum_{i=1}^{n}(\hat{Y}_i - \check{Y}_i)^2 \sum_{i=1}^{n}(\hat{Y}_i - \mathring{Y}_i)^2} \tag{27}$$

where n represents the number of data, Y is the actual values, $\hat{Y}$ the forecasted values, is the mean of the forecast values, and $\mathring{Y}$ is the mean of the real values (Zaranezhad et al., 2012).

In this research, all four mentioned criteria are used to evaluate estimates and forecast accuracy.

4. Research findings

4.1. Estimation and forecast with ARDL

In this section, the impact of the factors effective on carbon dioxide emissions is investigated in Iran. The carbon dioxide gas emission function is as follows:

$$CO2_t = f(CO2_{t-i}, GDP_{t-j}, M_{t-k})$$

$$CO2_t = \lambda_0 + \sum_{i=1}^{m}\lambda_{1i}CO2_{t-i} + \sum_{j=0}^{n}\lambda_{2j}GDP_{t-j} + \sum_{k=0}^{p}\lambda_{3k}M_{t-k} + \xi_t \tag{28}$$

where $CO2_t$ represents carbon dioxide emissions per millions of metric tons, $CO2_{t-I}$ is the carbon dioxide emissions in previous years, GDP_{t-j} is the country's GDP at constant prices in 2004 per billion Rials, M_{t-K} is the final energy consumption in terms of million barrels of crude oil equivalent, and it is time in years and I, j, and k represent lags of each of the exogenous regressors. The variables are used logarithmically for estimation. Using the logarithmic form of variables often reduces the variance of the heterogeneity and the coefficient of variables in the equation shows elasticities. The estimated carbon dioxide emission function is expressed as follows:

$$LCO2_t\text{: ARDL}(3,0,1)$$

$$\begin{aligned} LCO2_t &= -1.968 + 0.4372 LCO2_{t-1} + 0.0506 LCO2_{t-2} \\ &+ 0.5037 LCO2_{t-3} + 0.2104 LGDP_t + 0.388 LM_t - 0.5236 LM_{t-1} + \xi_t \end{aligned} \tag{29}$$

The number of optimal lags has been selected by the Schwartz–Bayesian criterion. The results of the ARDL dynamic model estimation are presented in Table 1.

Table 1. Results of dynamic ARDL model estimation

Variables and statistics	Coefficients
$LCO2_{t-1}$	.4372 (.011)
$LCO2_{t-2}$	.0506 (.736)
$LCO2_{t-3}$	.5037 (0)
$LGDP_t$	.2104 (.047)
LM_t	.388 (.038)
LM_{t-1}	-.5236 (.004)
R^2	.9969
$\bar{R}^2$	.996
F-probability	0
LM- probability	.135

Note: The numbers in parentheses show the probability of t.

Source: Authors computation from Microfit 4.1.

As can be seen in Table 1, the probability of t for all variables except carbon dioxide emissions with two lags is less than 0.05. Also, the H0 hypotheses that there is no relationship between explanatory, and thus dependent variables are rejected. As a result, carbon dioxide emissions show a positive and significant relationship with a lag with GDP and energy consumption. The absolute magnitude of the GDP coefficients and the final energy consumption are less than 1, indicating that the carbon dioxide gas emission ratio has a low elasticity toward GDP and energy consumption.

The value of the coefficient of determination and the adjusted coefficient of determination is more than 0.99, suggesting the high explanatory power of the regression variables. Moreover, F probability is smaller than 0.05, which indicates that the regression variables are generally significant. Given the fact that there is a dependent variable with lag on the right of the equation, instead of the Durbin-Watson value, the Durbin–Watson h probability is considered. The disadvantage of this test is that in the calculation formula of the Durbin–Watson h, the term under radical becomes negative and the Durbin–Watson h statistics cannot be calculated anymore; in such cases, alternate tests should be used. Since, in our study, the term under radical is negative and the Durbin–Watson h cannot be calculated, the Breusch–Godfrey LM test was used to check the correlation of the error terms. In this estimate, the probability of LM is greater than 0.05, suggesting that the H0 hypothesis is not rejected and that there is no serial correlation between residual terms.

Now, the existence of a long-term relationship between variables is investigated through the number of computational statistics of Banerjee et al. (1993). The results of the investigating the existence of a long-run relationship between variables at the 95% confidence level are presented in Table 2.

Table 2. The results of examining the existence of a long-run relationship at 95% confidence level

t computational statistics	Critical value k = 2	Existence of a long-term relationship
−.0208	−3.57	rejected

Note: K = Number of explanatory variables.

Source: Authors computation from Microfit 4.1.

According to Table 2, the absolute magnitude of the value of the Banerjee et al. (1993) statistics is greater than the absolute value of the computational value at 95% level of confidence. Therefore, according to the results, the existence of a long-run relationship between the variables is rejected.

The forecast results of carbon dioxide gas emission logarithm for 2011-2014 using ARDL model are presented in Table 3.

4.2 Forecasting with FLR

The prediction is based on the fuzzy linear regression model using the explanatory variables of GDP and marginal energy consumption. Fuzzy coefficients are obtained through linear programming and WinQSB software. After obtaining the coefficients, the final pattern is as follows:

$$L\tilde{\tilde{C}}O2_t = (0, 0.0844) + (0, 0)LGDP_t + (0.9021, 0)LM_t \tag{30}$$

4.3. Forecast with FARDL

In modelling using FARDL pattern, ARDL pattern components are used in the FLR model. In this model, the explanatory variables of carbon dioxide gas emissions are used with a lag, GDP, and marginal energy consumption. After obtaining the fuzzy coefficients through the WinQSB software, the model is as follows:

$$\begin{aligned} L\tilde{\tilde{C}}O2_t &= (0,0) + (0.2, 0.0151)LCO2_{t-1} + (0,0)LGDP_t \\ &+ (0.7235, 0)LM_t + \varepsilon_t \end{aligned} \tag{31}$$

The actual and forecast values (lower and upper limits) of the logarithms of carbon using various patterns are presented in Table 3.

Table 3. Results from of the forecast of carbon dioxide gas emission logarithm

Year	Actual value	Forecast and defuzzified value		
		ARDL	FLR	FARDL
2011	6.382423	6.3996	6.284136	6.300539
2012	6.392375	6.4494	6.283625	6.339928
2013	6.428074	6.4893	6.320495	6.380994
2014	6.470954	6.5458	6.365842	6.439837

Source: Author's computation from Microfit 4.1,WinQSB 2, and Excel 2010.

4.4. Comparing forecast accuracy of the model

After forecasting the logarithm of carbon dioxide emissions during the study period (2011-2015) using different patterns, the prediction accuracy of the patterns is compared using different criteria for prediction accuracy estimation. The results are presented in Table 4.

Table 4. Results of the evaluation of model forecast accuracy for carbon dioxide gas emission logarithms (million metric tons)

Model	RMSE	MAE	MAPE	R2
ARDL	0.0567	0.0525	0.8175	0.9444
FLR	0.1050	0.1049	1.6347	0.9869
FARDL	0.0322	0.0254	0.3982	0.9431

Source: Authors computation from Microfit 4.1,WinQSB 2,Excel 2010.

As shown in Table 4, the FARDL model has a lower RMSE, MAE, and MAPE than other models. Thus, it has a lower error and more precision and is suitable for predicting carbon dioxide emissions.

5. Conclusions

5.1. Summary of findings

After evaluating the forecast accuracy of different patterns using four evaluation criteria, it was determined that the FARDL pattern has a lower RMSE, MAE, and MAPE than other patterns, resulting in less error and more accuracy. Thus, it is appropriate to predict carbon dioxide emissions.

This study showed that carbon dioxide gas emissions with a lag, GDP, and energy consumption have a positive and significant correlation with carbon dioxide gas emission In addition; the absolute value of the GDP and marginal energy consumption is less than 1, suggesting that the carbon dioxide emission rate has a low elasticity towards GDP and energy consumption.

The government can save energy and reduce energy consumption by the optimised pricing of different energy sources in different parts of the country. Energy price in Iran is lower in all sectors than in many countries. Hence, it leads to an increase in energy consumption in Iran. The high energy consumption in Iran has also led to a higher rate of an increase in carbon dioxide emissions than GDP. Regarding the pricing policies in different sectors, adjusting the optimum prices is a sensitive issue. Energy prices should not be too high to reduce production, and should not be so low that producers have no incentive to save. For example, in some countries, household electricity prices are higher than for the industrial sector, which is as an incentive for producers. Furthermore, it causes that household customers have better electricity management. In contrast, in Iran according to Tavanir's (s.a.) statistics in 2014, the price of electricity in the industrial sector was 23.1 times higher than the household sector. The price of electricity in the agricultural sector is negligible compared to the rest of the sector so that the price of electricity in the industrial sector is more than triple in agriculture in 2014. Such a pricing policy reduces the incentive for the agricultural sector to lower energy consumption and saving. In the household sector, rising electricity prices can make people more sensitive to electricity consumption, so adjusting energy prices across different parts of the country and approaching real and reasonable prices can save more and reduce energy consumption in all sectors. As a result, the costs for the country and environmental pollution would be reduced.

Among other factors that can reduce carbon dioxide emissions and, thereby, reduce environmental pollution one can name strict environmental regulations, producing and importing machinery and equipment that meet the required standards.

The limitations of this research were the lack of access to more data, while with more data; the accuracy of prediction would increase. Researchers can use this model (FARDL) in future papers to predict other variables.

5.2. Research contribution

Several studies abound in the area of estimation of carbon dioxide emission in Iran but in this study, the main goal was to create a new hybrid model and compare the accuracy of prediction of the hybrid model with the base models. Tseng et al. (2001) model is more similar to this model. Their hybrid model also has more predictive accuracy than base models.

5.3. Research implication

Most models require a lot of data to make accurate prediction, but sometimes there is not much data available, or data collection requires a lot of time and cost. In this study the main purpose was to create a hybrid model that, even with low data, has a more accurate prediction than base models, and therefore, only two variables of gross domestic product and energy consumption that have the greatest impact on the emission of carbon dioxides are used to estimate and forecast. As in the research by Tseng et al. (2001), here also the hybrid model has increased predictive accuracy compared to base models. This hybrid model, like the FARIMA hybrid model, is also applicable to other countries and other variables.

5.4. Research limitations and future works

Researchers can use this hybrid model for accurate prediction when data is low or other models are difficult. In addition, researchers can compare this model's prediction accuracy with other models in the future. The limitation of this study was the lack of a similar hybrid model for comparison.

References

Abbasinejad, H., & Goodarzi Farahani, Y. (2013). *Eghtesad sanjie karbordi ba narm afzarhaye Microfit va Eviews* [Applied econometric with Microfit and Eviews softwares] (1st ed.). Tehran: Noore Elm Publication. (in Persian)

Acaravci, A., & Ozturk, I. (2010, December). On the relationship between energy consumption, CO_2 emissions and economic growth in Europe. *Energy, 35*(12), 5412-5420. doi: 10.1016/j.energy.2010.07.009

Amiri, M. (2010). Group selection for the selection of machine tools with using fuzzy VIKOR method. *Journal of Industrial Management Studies, 6*(16), 167-188. doi: JR_JIMS-6-16_007

Banerjee, A., Dolado, J. J., & Mestre, R. (1993). *On some simple tests for cointegration: The cost of simplicity*. Madrid: Banco de España, Servicio de Estudios.

Dubois, D., & Prade, H. (1980). *Fuzzy sets and systems: Theory and applications* (Mathematics in Science & Engineering, Vol. 114). New York: Academic Press.

Energy Information Administration [EIA]. (2016). *Total Carbon Dioxide Emissions from the Consumption of Energy*. Washington, DC: Author. Retrieved from https://www.eia.gov/beta/international/data/browser/#/?pa=0000000000000000000000000000000000000000000000000000001&c=00000000000000000000002&ct=0&tl_id=40-A&vs=INTL.4008-8-IRN-MMTCD.A&ord=CR&cy=2009&vo=0&v=H&end=2016

Fukami, S., Mizumoto, M., & Tanaka, K. (1980). Some considerations is fuzzy conditional inference. *Fuzzy Sets and Systems, 4*(3), 243-273. doi: 10.1016/0165-0114(80)90014-7

Hamit-Haggar, M. (2012). Greenhouse gas emissions, energy consumption and economic growth: A panel cointegration analysis from the Canadian industrial sector perspective. *Energy Economics, 34*(1), 358-364. doi: 10.1016/j.eneco.2011.06.005

Ishibuchi, H., & Tanaka, H. (1988). *Kongō 0-1 seisū keikaku mondai niyoru kukan kaiki bunseki* [Interval regression analysis based on mixed 0-1 integer programming problem]. *Journal of Japanese Industrial Management Association, 40*(5), 312-319. (in Japanese)

Jafari Samimi, A., & Mohammadi Khayareh, M. (2014). *Rabeteye kootah modat va boland modat beyne enteshare di okside karbon, masrafe enerji va roshde eghtesadi: shavahede jadid dar iran* [Short-term and long-term relationship between carbon dioxide emissions, energy consumption and economic growth: New evidence in Iran]. *Economic Researches, 14*(2), 1-20. (in Persian)

Johansen, S., & Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics, 53*(1-3), 211-244. doi: 10.1016/0304-4076(92)90086-7

Kazemi, A., & Hosseinzadeh, M. (2012). *Pishbinie taghazaye bargh dar bakhshhaye mokhtalefe Iran ba estefadeh az regression khatie fazi* [Forecasting electricity demand in different parts of Iran using fuzzy linear regression.] The 9th International Energy Conference, 2nd and 3rd of March, Tehran, Iran. (in Persian)

Lotfalipour, M., R. Fallahi, M. A., & Ashena, M. (2011). Investigating the relationship between carbon dioxide emissions and economic growth, energy and trade in Iran. *Journal of Economic Research, 46*(1), 151-173. doi: 10.1016/j.energy.2010.08.004

Ministry of Energy of Iran. (2010). *Energy Balance Sheet of Iran*. Tehran: Author. Retrieved from http://pep.moe.gov.ir/%D8%A7%D9%86%D8%AA%D8%B4%D8%A7%D8%B1%D8%A7%D8%AA/%D8%B3%D8%B1%DB%8C-%D8%B2%D9%85%D8%A7%D9%86%DB%8C/%D8%B3%D9%87%D9%85-%D9%88-%D8%B1%D8%B4%D8%AF-%D8%AF%D8%B1-%D8%AA%D8%B1%D8%A7%D8%B2%D9%86%D8%A7%D9%85%D9%87-%D8%A7%D9%86%D8%B1%DA%98%DB%8C

Moghaddasi, R., & Golriz Ziaei, Z. (2011). *Barrasie rabeteyeh beyne enteshare gaze di okside karbon va tolide nakhalese dakheli bar asase dadehaye tarkibi* [Investigating the relationship between carbon dioxide gas emissions and gross domestic product based on combined data]. *Economics and Agricultural Development Studies, 25*(4), 480-487. doi: 10.22067/jead2.v0i0.12187 (in Persian)

Mohammad Bagheri, A. (2010). *Barrasie ravabete kootah modat va boland modat beyne tolide nakhalese dakheli, masrafe enerji va enteshare di okside karbon dar Iran* [Study of short-run and long-run relationships between GDP, energy consumption and carbon dioxide emissions in Iran]. *Quarterly Journal of Energy Economics, 7*(27), 101-129.

Noferesti, M. (1999). *Rishe vahed va hamjam'i dar eghtesad* [Unit root & cointegration in econometrics] (1st ed.). Tehran: Khadamate Farhangi Resa. (in Persian)

Pahlavan, M., & Dahmardeh, N. (2007). Import and export demand functions estimated in the Iranian economy using ARDL convergence. *Journal of Economic, 3*, 110-120.

Pesaran, M. H., & Shin, Y. (1996, March-April). Cointegration and speed of convergence to equilibrium. *Journal of Econometrics, 71*(1-2), 117-143. doi: 10.1016/0304-4076(94)01697-6

Pesaran, M. H., & Shin, Y. (1998). Chapter 11 – An autoregressive distributed-lag modelling approach to cointegration analysis. In S. Strøm (Ed.), *Econometrics and economic theory in the 20th Century* (pp. 371-413). doi.: 10.1017/CCOL521633230.011

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16*, 289-326. doi: 10.1002/jae.616

Pesaran, M. H., & Smith, R. P. (1998). Structural analysis of cointegrating VARs. *Journal of Economic Surveys, 12(5)*, 471-505. doi: 10.1111/1467-6419.00065

Sadeghi, H., & Islami Ordoghly, M. (2011), *Roshde eghtesadi va aloodegi mohite zist dar keshvarhaye ozve peymane Kyoto* [Economic growth and environmental pollution in the Kyoto protocol parties]. *Quarterly Journal of Energy Economics, 30*(8), 1-32. (in Persian)

Sadeghi, K., & Ebrahimi, S. (2013). *Tasire tose'e mali, tolide nakhalese dakheli va masrafe enerji bar aloodegi mohite zist dar Iran (rahyafte ARDL)* [The effect of financial development, GDP and energy consumption on environmental pollution in Iran (ARDL approach)]. *Iranian Energy Economics (Environment and Energy), 7*(2), 43-73. (in Persian)

Soytas, U., & Sari, R. (2009). Energy consumption, economic growth, and carbon emissions: Challenges faced by an EU candidate member. *Ecological Economics, 68*(6), 1667-1675. doi: 10.1016/j.ecolecon.2007.06.014

Soytas, U., Sari, R., & Ewing, B. T. (2007). Energy consumption, income, and carbon emissions in the United States. *Ecological Economics, 62*(3-4), 482-489. doi: 10.1016/j.ecolecon.2006.07.009

Tanaka, H., & Ishibuchi, H. (1992). Possibilistic regression analysis based on Linear Programming. In J. Kacprzyk, M. Fedrizzi (Eds.), *Studies in Fuzziness. Fuzzy Regression Analysis* (Vol. 1, pp. 47-60). Warsaw: Omnitech Press.

Tanaka, H., Shimomura, T., Watada, J., & Asai, K. (1984). *Fuzzy linear regression analysis of the number of staff in local government*. FIP-84 at Kauai, Hawaii, 22-26.

Tavanir. (s.a.). Retrieved from http://amar.tavanir.org.ir/

Tseng, F. M., Tzeng, G. H., Yu, H. C., & Yuan, B. J. (2001, February). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and Systems, 118*(1), 9-19. doi: 10.1016/S0165-0114(98)00286-3

Yen, K. K., Ghoshray, S., & Roig, G. (1999). A linear regression model using triangular fuzzy number coefficients. *Fuzzy Sets and Systems, 106*(2), 167-177. doi: 10.1016/S0165-0114(97)00269-8

Zadeh, L. A. (1965, June). Fuzzy sets. *Information and Control, 8*(3), 338-353. doi: 10.1016/S0019-9958(65)90241-X

Zaranezhad, M., & Saadatmehr, M. (2007). *Takhmineh taghaza barayeh goosht dar Iran* [Estimate the demand for meat in Iran]. *Journal of Humanities and Social Sciences (Particularly Economics), 7.1*(26), 63-82. (in Persian)

Zaranezhad, M., Ebrahimi, S., Raoofi, A., & Kiani, P. (2012). *Pishbinie gheymate nafte khame OPEC ba estefadeh az modele khod regresioni miyangine motehareke anbashteye fazi* [OPEC crude oil price forecast using the fuzzy stacked moving average recursive model]. *Quarterly Journal of Energy Economics of Iran, 2(5),* 107-207. (in Persian)

Zimmermann, H. J. (1996). Fuzzy control. In H. J. Zimmerman, *Fuzzy set theory – and its applications* (pp. 223-264). Netherlands: Springer. doi: 10.1007/978-94-010-0646-0