Applied Soft Computing 9 (2009) 590–598

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A novel fuzzy linear regression model based on a non-equality possibility index and optimum uncertainty

H. Shakouri G. *, R. Nadimi

Industrial Engineering Department, University of Tehran, Tehran, Iran

ARTICLE INFO

Article history:
Received 7 August 2007
Received in revised form 7 July 2008
Accepted 10 August 2008
Available online 29 August 2008

Keywords:
Fuzzy linear regression
Fuzzy equality
Non-equality possibility
Uncertainty

ABSTRACT

Various kinds of fuzzy regression models are introduced in the literature and many different methods are proposed to estimate fuzzy parameters of the models. In this study, a new approach is introduced to find the parameters of a linear fuzzy regression, with fuzzy outputs, the input data of which is measured by crisp numbers. Based on a non-equality possibility index, a new objective function is designed and solved, by which a minimum degree of acceptable uncertainty (the h-level or h-cut) is found. Four numerical examples are presented to compare the proposed approach with some other methods. Results show superiority of the new approach based on the criterion used by Kim and Bishu in the cases studied here. A realistic application of the proposed method is also presented, by which the total energy consumption of the Residential-Commercial sector in Iran is modeled using three variables of the GDP, number of the Households and an Energy Price index as inputs (exogenous variables) to the model.

1. Introduction

A fuzzy regression (FR) is an input–output relation in which the input or the output, or both are fuzzy numbers. Unlike the classic linear regression in which the parameters are assumed to be random variables with probability distribution functions, in a fuzzy regression, the coefficients are subject to the *possibility theory* [1]. Therefore, the input data (independent variable, X), the output data (dependent variable Y) and consequently the relationship between them are relaxed.

A general form of the linear fuzzy regression (LFR) is represented by

$$\tilde{Y}_i^* = \tilde{A}_0\tilde{X}_{i0} + \tilde{A}_1\tilde{X}_{i1} + \cdots + \tilde{A}_n\tilde{X}_{in} \tag{1}$$

or in a vector production form:

$$\tilde{Y}_i^* = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{X}}_i \tag{2}$$

where $\tilde{Y}_i^*$ and $\tilde{X}_i$ are the fuzzy output and vector of the input observations, respectively. Hence, $\tilde{\boldsymbol{A}} = [\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_n]$ and $\tilde{\boldsymbol{X}}_i = [\tilde{X}_{i0}, \tilde{X}_{i1}, \ldots, \tilde{X}_{i,n}]^t$; $i = 1, \ldots, m$; so that, $\tilde{A}_j$, $j = 0, 1, \ldots, n$ are the fuzzy coefficients.

The fuzzy regression analysis is a powerful tool for investigating and predicting data sets by measuring a vague concept that contains a degree of ambiguity, uncertainty or *fuzziness* [2–5]. The main purpose of fuzzy regression models is to find the best model with the least error. Depending on how we define the error, this method can be classified into two classes:

(1) *Possibilistic approach*, which tries to minimize the whole fuzziness of the model by minimizing the total spreads of its fuzzy coefficients, subject to including the data points of each sample within a specified feasible data interval [6–8].

The fuzzy regression analysis was first introduced by Tanaka et al. [6], who established his idea on the basis of the possibility theory. He modeled the procedure of parameter estimation as a linear programming problem, where the inputs are crisp and the output is a fuzzy number. Later, the optimization model was solved subject to that the observations fall in the fuzzy sets computed by the model. He then extended triangular fuzzy coefficients to Gaussian fuzzy numbers [9]. Some discussions have been presented offering some modifications on the solution of the so-called exponential possibility regression problems, especially on determining the center of the possibility distribution [10], where it was shown that the estimated parameters will be quite different if the main nonlinear problem is solved approximately by dividing it into two linear programming problems.

Recently, Ge and Wang [12] tried to determine the relationship between threshold value (h-parameter) and input data when data contains a considerable level of noise or uncertainty. They used the threshold value to measure degrees of fitness in fuzzy linear regression. Eventually, they showed

* Corresponding author.
E-mail address: hshakouri@ut.ac.ir (H. Shakouri G.).

1568-4946/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2008.08.005

that the parameter h is inversely proportional to the input noise. Meantime, many researchers recommended a combination of fuzzy regression models with some other approaches, like Neuro-Fuzzy modeling [2], TSK-FR modeling [5] and Monte-Carlo methods [13] to improve the result obtained from ordinary LFR.

(2) *Least squares model*, which minimizes the sum of squared errors in the estimated value, based on their specifications [11–21].

This approach is indeed a fuzzy extension of the ordinary least squares, which obtains the best fitting to the data, based on the distance measure under fuzzy consideration, applying information included in the input–output data set.

In the present study, the possibilistic approach is employed, for which a new objective function is introduced. The main idea in this study, which is derived from the second class of the fuzzy regressions, is to minimize the distance between the outputs of the model and the measurements. The proposed objective function helps to estimate an optimal confidence level, namely h-level, simultaneously with the parameters.

The organization of the remaining parts of the paper is as follows: in Sections 2 and 3, preliminary definitions, including fuzzy numbers and fuzzy linear regression models are given. Furthermore, some important remarks in these models are described. In Section 4, the new approach is introduced, based on an index proposed to measure inequality between two fuzzy numbers. Numerical examples are applied to compare the results of our approach with that of some existing methods in Section 5, along with some criteria that are conducted to assess the performance of the methods. Finally, a realistic application is given in Section 6, and Section 7 concludes the paper.

2. Fuzzy numbers

Three introductory definitions are presented in this section. First, the definition for the fuzzy numbers used through this paper is given based on the concept defined by Dubois and Prade [22]:

Definition 1. $\tilde{A}$ is a fuzzy number of the LR-type if there are a, $c_L > 0$, $c_R > 0$ in $\mathbf{R}$ so that:

$$\mu_{\tilde{A}}(x) = \begin{cases} L\left(\dfrac{a-x}{c_L}\right) & \text{for } x \le a \\ R\left(\dfrac{x-a}{c_R}\right) & \text{for } x \ge a \end{cases} \tag{3}$$

where L and R are decreasing functions from $\mathbf{R}^+$ to [0,1], and $L(x) = R(x) = 1$, for $x \le 0$, and $L(x) = R(x) = 0$ for $x \ge 1$. If $L(x) = R(x) = 1 - x$, then the following notation represents a triangular fuzzy numbers:

$$\tilde{A} = (c_L,\ a,\ c_R)_{\mathrm{LR}} \tag{4}$$

where a is called the mean value or the *center*, and c_L and c_R are called the left and right *spreads*, respectively. Furthermore, if $c_L = c_R = c$ then $\tilde{A}$ is called a symmetric triangular fuzzy number, denoted by

$$\tilde{A} = (a,\ c)_L \tag{5}$$

Definition 2. By applying the Extension Principle [1], for any function of several fuzzy numbers, e.g. the fuzzy linear regression model, the membership function of $\tilde{Y}_i^*$ in (1), can be defined by

$$\mu_{\tilde{Y}_i^*}(\nu) = \operatorname*{Sup}_{y_i^* = f(X_i, \tilde{A})} \cap \mu_{\tilde{A}_j}(\alpha) \tag{6}$$

where ν and α are dummy variables, $\cap$ stands for the t-norm, $\tilde{A} = (\boldsymbol{a}, \mathbf{c})$, $\mathbf{a} = [a_0, \ldots, a_n]$, $\mathbf{c} = [c_0, \ldots, c_n]$ and $\mathbf{X}_i$'s are non-fuzzy ordinary variables, like the inputs of a fuzzy regression with crisp inputs, which are arranged in an n-vector: $\mathbf{X}_i = [X_{i0}, \ldots, X_{in}]^t$, $i = 1, \ldots, m$. Based on this principle, the fuzziness of the coefficients in (1) is extended to the output of the regression model.

Definition 3. In order to define the fuzzy equality between fuzzy numbers, Dubois and Prade [22] proposed the following index:

$$Poss(\tilde{A}_1 = \tilde{A}_2) = \operatorname*{Sup}_{x \in R^1} \min\{\mu_{\tilde{A}_1}(x), \mu_{\tilde{A}_2}(x)\} \tag{7}$$

where *Poss* is short for Possibility.

In this paper, for simplicity purposes, we consider only triangular fuzzy numbers.

3. Fuzzy linear regression models

The fuzzy regression analysis introduced by Tanaka et al. [6], considers crisp inputs and fuzzy outputs. The general model given by (1) is then represented as follows:

$$\tilde{Y}_i^* = \tilde{A}_0 X_{i0} + \tilde{A}_1 X_{i1} + \cdots + \tilde{A}_n X_{in} = \tilde{\boldsymbol{A}}\boldsymbol{X}_i \tag{8}$$

where $\tilde{Y}_i^*$ is the fuzzy output, $\tilde{A}_j$, $j = 0, 1, \ldots, n$; are the fuzzy coefficients, and $\mathbf{X}_i$ is the vector of crisp inputs. The optimization process is formulated as follows:

$$\operatorname{Min} Z(h) = \sum_{i=1}^{m}\sum_{j=0}^{n} c_j |x_{ij}|$$

subject to

$$\begin{aligned} &\sum_{j=0}^{n} a_j x_{ij} + |L^{-1}(h)| \sum_{j=0}^{n} c_j |x_{ij}| \ge y_i + |L^{-1}(h)| e_i, \quad i = 1, 2, \ldots, m \\ &\sum_{j=0}^{n} a_j x_{ij} - |L^{-1}(h)| \sum_{j=0}^{n} c_j |x_{ij}| \le y_i - |L^{-1}(h)| e_i, \quad i = 1, 2, \ldots, m \\ &c_j \ge 0, a_j \in \boldsymbol{R} \quad j = 0, 1, \ldots, n. \end{aligned} \tag{9}$$

where $\mathbf{a} = [a_0, a_1, \ldots, a_n]$ and $\mathbf{c} = [c_0, c_1, \ldots, c_n]$ for $a_j, c_j \in \mathbf{R}$ are the center and spread of the estimated coefficients, $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_m] = \{x_{ij}\}$ is the observation matrix of the inputs, and $\tilde{Y}_i = (y_i, e_i)$; $i = 1, \ldots, m$, are the fuzzy output observations. If the coefficients are assumed to be triangular fuzzy numbers, then $|L^{-1}(h)|$ is supposed to be equal to $|L(h)| = 1 - h$; $0 < h < 1$. In fact, h is a confidence level; thereupon one idea to narrow down the fuzzy numbers is to apply higher h-levels. Consequently, from a pessimistic point of view, the h-level is close to zero, while from an optimistic one, the h-level will be close to one. Later, Chang and Lee [23] studied the constraint $c_j \ge 0$, $j = 1, \ldots, n$, and replaced it by $\mathbf{c}|\mathbf{x}_i| \ge 0$; $i = 1, \ldots, m$.

4. The new approach

There are three types of the possibilistic linear regression analysis, so-called the *Min*, *Max* and *Conjunction* problems. Based on the notation used by Tanaka at el. [7], suppose that $\bar{A}_j$, $\underline{A}_j$ and $\hat{A}_j$; $j = 1, \ldots, n$, are estimated such that we have one of the following cases

$$\tilde{Y}_i \subseteq_h \bar{Y}_i \quad (\text{the } Min \text{ problem}) \tag{10}$$

$$\tilde{Y}_i \supseteq_h \underline{Y}_i \quad (\text{the } Max \text{ problem}) \tag{11}$$

$$[\tilde{Y}_i]_h \cap [\hat{Y}_i]_h \ne \quad (\text{the } Conjunction \text{ problem}) \tag{12}$$