Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 274 (2015) 97–108

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy linear regression using rank transform method

Hye-Young Jung [a], Jin Hee Yoon [b], Seung Hoe Choi [c,*]

[a] Dept. of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, South Korea
[b] School of Mathematics and Statistics, Sejong University, Neungdong-ro, Gwangjin-gu, Seoul, 143-747, South Korea
[c] School of Liberal Arts and Science, Korea Aerospace University, Koyang, 411, South Korea

Received 6 June 2013; received in revised form 16 October 2014; accepted 5 November 2014

Available online 11 November 2014

Abstract

In regression analysis, the rank transform (RT) method is known to be neither dependent on the shape of the error distribution nor sensitive to outliers. In this paper, we construct a so-called α-level fuzzy regression model based on the resolution identity theorem and apply RT method to this model. Fuzzy regression models with crisp input/fuzzy output and fuzzy input/fuzzy output are investigated to show the effectiveness of the proposed method. To compare its effectiveness with existing methods, we introduce a new performance measure. In addition, we propose a method to obtain a predicted output with respect to a specific target value and show that our model is more robust compared with other methods when the data contain some outliers.

1. Introduction

The fuzzy regression model introduced by Tanaka et al. [32] can be classified into two types based on the functional relationship between dependent and independent variables. If the functional relationship is known, the model is called a parametric fuzzy regression model, otherwise, it is called a nonparametric fuzzy regression model. Many methods have been proposed to construct parametric and nonparametric fuzzy regression models.

These methods are classified into numerical and statistical methods. Numerical methods identify the fuzzy regression model by minimizing the sum of the spreads of the estimated dependent variable. Many authors have constructed fuzzy regression models using numerical methods such as linear or nonlinear programming [10,20,21,25,28,31,32]. In other studies, various statistical methods have been suggested to construct fuzzy regression models. Some of them have used the least squares method [3,7,9,18,22,24,37]. Choi and Buckley [6] and Taheri and Kelkinnama [30] suggested the least absolute deviations method, which is an alternative to the method of least squares. Generally, fuzzy regression analysis has been criticized because it is sensitive to outliers, also the spreads of the estimated value become wider as more data are included in the model [6,26]. Nonparametric methods [5,21,34] have been suggested to overcome these drawbacks. The rank transform (RT) method is one of the nonparametric methods known to be

* Corresponding author. Tel.: +82 2 300 0073; fax: +82 2 300 0492.
E-mail address: shchoi@kau.ac.kr (S.H. Choi).

http://dx.doi.org/10.1016/j.fss.2014.11.004

neither dependent on the shape of the error distribution nor sensitive to outliers. Iman and Conover [16,17] showed that the RT method is a robust and powerful procedure in hypothesis testing with respect to experimental designs. Since then, the RT method has been widely used in multiple regression analysis and ANOVA [2,13]. Some authors applied an RT method to analyze their data [23,35], and others showed statistical properties such as consistency of RT estimators [33,40]. This paper applies the rank transform (RT) method to fuzzy regression model, and we confirm that the proposed method is a robust method which is not sensitive to fuzzy outliers through examples. Because detecting and handling outliers is important in data analysis, many studies [4,15,27] have been carried out to deal with outliers.

In this paper, we introduce an RT method that uses the rank of the modes and endpoints of the α-level sets of fuzzy numbers for the purpose of developing a fuzzy regression model. In addition, we investigate a method to obtain a predicted output with respect to a specific target value in Section 4. Then we compare the efficiency of the proposed regression model with methods based on least squares and absolute deviations using some data which contain fuzzy outliers.

2. Fuzzy regression model

In order to explain the functional relationship among incompletely informed variables, we consider the fuzzy regression model that can be expressed as follows:

$$Y_i(\mathbf{X}_i) = A_0 + A_1 X_{i1} + \cdots + A_p X_{ip}, \tag{1}$$

where X_{ij}, A_j, and $Y_i(\mathbf{X}_i)$ are LR-fuzzy numbers [11].

One of the purposes of fuzzy regression analysis is to determine the regression coefficients that minimize the difference between the observed fuzzy numbers and predicted fuzzy numbers based on the observed data $\{(X_{i1}, \cdots, X_{ip}, Y_i) : i = 1, \cdots, n\}$.

The membership function of the LR-fuzzy number $A = (a, l_a, r_a)_{LR}$ is

$$\mu_A(x) = \begin{cases} L_A((a-x)/l_a) & \text{if } 0 \le a - x \le l_a, \\ R_A((x-a)/r_a) & \text{if } 0 \le x - a \le r_a, \\ 0 & \text{otherwise,} \end{cases}$$

where L_A and R_A are monotonic decreasing functions that satisfy $L_A(0) = R_A(0) = 1$ and $L_A(1) = R_A(1) = 0$. Here, a denotes the mode of the fuzzy number A, and l_a and r_a denote the left and right spreads of the fuzzy number A, respectively. If $L_A(x) = R_A(x) = 1 - x$, then the LR-fuzzy number A is called a triangular fuzzy number and is represented as $(a, l_a, r_a)_T$. In particular, we express the LR-fuzzy number as $(a, s_a)_{LR}$, when the fuzzy number is symmetric, that is, the left and right spreads are identical.

The α-level set of the LR-fuzzy number $A = (a, l_a, r_a)_{LR}$ is defined as

$$A(\alpha) = \begin{cases} \overline{\{x : \mu_A(x) > \alpha\}} & \text{if } \alpha = 0 \\ \{x : \mu_A(x) \ge \alpha\} & \text{if } 0 < \alpha \le 1, \end{cases}$$

where $\bar{A}$ denotes the closure of A. The α-level set of the fuzzy number A is the closed interval with mode a, left spread $(l_a L_A^{-1}(\alpha))$, and right spread $(r_a R_A^{-1}(\alpha))$. Hence, we can represent the α-level set of the fuzzy number A as follows:

$$A(\alpha) \doteq \left[a - l_a L_A^{-1}(\alpha), a + r_a R_A^{-1}(\alpha)\right].$$

Thus, the α-level set of the observed fuzzy number $Y_i = (y_i, l_{y_i}, r_{y_i})_{LR}$ is

$$Y_i(\alpha) \doteq \left[y_i - l_{y_i} L_{Y_i}^{-1}(\alpha), y_i + r_{y_i} R_{Y_i}^{-1}(\alpha)\right]$$

and the α-level set of the predicted fuzzy numbers $Y_i(\mathbf{X}_i)$ is

$$\sum_{k=0}^{p} \left[l_{A_k}(\alpha), r_{A_k}(\alpha)\right] \cdot \left[l_{X_{ik}}(\alpha), r_{X_{ik}}(\alpha)\right], \tag{2}$$