Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Fuzzy Sets and Systems 135 (2003) 305–316

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy least-squares algorithms for interactive fuzzy linear regression models

Miin-Shen Yang*, Hsien-Hsiung Liu

Department of Mathematics, Chung Yuan Christian University, Chung-Li, Taiwan 32023, Republic of China

Received 10 August 2001; accepted 20 March 2002

Abstract

Fuzzy regression analysis can be thought of as a fuzzy variation of classical regression analysis. It has been widely studied and applied in diverse areas. In general, the analysis of fuzzy regression models can be roughly divided into two categories. The first is based on Tanaka's linear-programming approach. The second category is based on the fuzzy least-squares approach. In this paper, new types of fuzzy least-squares algorithms with a noise cluster for interactive fuzzy linear regression models are proposed. These algorithms are robust for the estimation of fuzzy linear regression models, especially when outliers are present. Numerical examples are given to detail the effectiveness of this approach. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Fuzzy sets; Regression models; Estimation; Fuzzy least squares; Linear programming; Noise cluster; Outlier

1. Introduction

Regression analysis is used to model the functional relationship between dependent and independent variables. In conventional regression analysis, deviations between the observed values and the estimates are assumed to be due to random errors. Thus, statistical techniques are applied to perform estimation and inference in regression analysis. However, the deviations are sometimes due to the indefiniteness of the structure of the system or imprecise observations. The uncertainty in this type of regression model becomes fuzziness, not randomness. Since Zadeh [18] proposed fuzzy sets, fuzziness has received more attention. Now fuzzy data analysis has become increasingly important (see [2]).

Tanaka et al. [14] first proposed a study in linear regression analysis with a fuzzy model. They considered the parameter estimations of fuzzy linear regression (FLR) models under two factors,

* Corresponding author. Tel.: +886-3-456-3171; fax: +886-3-456-3160.
E-mail address: msyang@math.cycu.edu.tw (Miin-Shen Yang).

PII: S0165-0114(02)00123-9

namely the degree of the fitting and the vagueness of the model. The estimation problems were then transformed into linear programming (LP) based on these two factors. This type of analysis of FLR models is called Tanaka's approach. The extension of FLR models and different estimation methods have been proposed by many researchers. Since the measure of best fitting by residuals under fuzzy consideration is not presented in Tanaka's approach, Diamond [6] proposed the so-called fuzzy least-squares approach, which is a fuzzy extension of the ordinary least squares based on a new defined distance on the space of fuzzy numbers. According to Zadeh's construction of fuzzy sets as a basis for a theory of possibility [19], fuzzy regression analysis is also named as a possibility regression analysis. Thus, Tanaka's approach to possibility regression analysis, instead of the measure of best fitting by residuals, uses linear programming inclusion relations. However, the fuzzy least-squares approach to possibility regression analysis, does not consider inclusion relations, directly uses the best fitting measure by residuals and information included in the input–output data under fuzzy consideration. Fuzzy regression models and estimation techniques have been widely studied and applied in diverse areas (see [1,4,9–11]). Generally, these fuzzy regression methods can be roughly divided into two categories. The first is based on Tanaka's LP approach (see [9–14]). The second category is based on the fuzzy least-squares approach (see [1,6,16–17]).

A fuzzy number A is defined as a convex normalized fuzzy set of the real line $\mathbb{R}$ so that there exists exactly one $x_0 \in \mathbb{R}$ with $\mu_A(x_0) = 1$, and its membership $\mu_A(x)$ is piecewise continuous. A fuzzy number M is of the LR-type if there are $m, \alpha > 0, \beta > 0$ in $\mathbb{R}$ so that

$$\mu_M(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right) & \text{if } x \leqslant m, \\ R\left(\dfrac{x-m}{\beta}\right) & \text{if } x \geqslant m, \end{cases}$$

where L and R are decreasing functions from $\mathbb{R}^+$ to $[0,1]$, and $L(x) = R(x) = 1$, for $x \leqslant 0$; 0 for $x \geqslant 1$. m is called the center value of M and α and β are called the left and right spreads, respectively. Symbolically, M is denoted by $M = (m, \alpha, \beta)_{LR}$ (see [20]). Let M and N be two LR-type fuzzy numbers with $M = (m, \alpha, \beta)_{LR}$ and $N = (n, \gamma, \delta)_{LR}$. Then, by the extension principle, the following operations are defined: $M + N = (m+n, \alpha+\gamma, \beta+\delta)_{LR}$; $-N = (-n, \delta, \gamma)_{RL}$; $(m, \alpha, \beta)_{LR} - (n, \gamma, \delta)_{RL} = (m - n, \alpha + \delta, \beta + \gamma)_{LR}$; $\lambda(m, \alpha, \beta)_{LR} = (\lambda m, \lambda\alpha, \lambda\beta)_{LR}$ when $\lambda > 0$, $\lambda(m, \alpha, \beta)_{LR} = (\lambda m, -\lambda\beta, -\lambda\alpha)_{RL}$ when $\lambda < 0$ (see [7]). For an LR-type fuzzy number $A = (a, \alpha, \beta)_{LR}$, if L and R are of the form $T(x) = 1 - x$ for $0 \leqslant x \leqslant 1$ and 0 otherwise, A is called a triangular fuzzy number, denoted by $A = (a, \alpha, \beta)_T$. If $\alpha = \beta$, $A = (a, \alpha, \alpha)_T$ is called a symmetrical triangular fuzzy number, denoted by $A = (a, \alpha)_T$.

Tanaka et al. [14] considered the following FLR model:

$$Y^* = A_0 + A_1 x_1 + \cdots + A_p x_p,$$

where $\underline{x}' = (x_0, x_1, \ldots, x_p)$ are non-fuzzy inputs (with $x_0 = 1$) and $A_0, A_1, \ldots, A_p$ are symmetrical triangular fuzzy parameters with $A_i = (a_i, \alpha_i)_T$, $\alpha_i \geqslant 0$, $i = 0, 1, \ldots, p$, in which the model is also called as a non-interactive FLR. Then $Y^* = A_0 x_0 + A_1 x_1 + \cdots + A_p x_p = (\sum_{i=0}^{p} a_i x_i, \sum_{i=0}^{p} \alpha_i |x_i|)_T$ is also a symmetrical triangular fuzzy output. Let $\{(\underline{x}_j, Y_j),\ j = 1, \ldots, n\}$ be a data set with $Y_j = (y_j, e_j)_T$, $j = 1, \ldots, n$. The different parameter estimation methods a_i, α_i, $i = 1, \ldots, p$ were studied in Tanaka's approach [11–14] and also in fuzzy-least squares approach [6,16–17].