ELSEVIER

Fuzzy Sets and Systems 126 (2002) 389–399

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy least-squares linear regression analysis for fuzzy input–output data

Miin-Shen Yang *, Tzu-Shun Lin

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

Received 27 July 1997; received in revised form 7 September 2000; accepted 12 February 2001

Abstract

A fuzzy regression model is used in evaluating the functional relationship between the dependent and independent variables in a fuzzy environment. Most fuzzy regression models are considered to be fuzzy outputs and parameters but non-fuzzy (crisp) inputs. In general, there are two approaches in the analysis of fuzzy regression models: linear-programming-based methods and fuzzy least-squares methods. In 1992, Sakawa and Yano considered fuzzy linear regression models with fuzzy outputs, fuzzy parameters and also fuzzy inputs. They formulated multiobjective programming methods for the model estimation along with a linear-programming-based approach. In this paper, two estimation methods along with a fuzzy least-squares approach are proposed. These proposed methods can be effectively used for the parameter estimation. Comparisons are also made between them.

Keywords: Fuzzy linear regression; Estimation; Multiobjective programming; Fuzzy least-squares; Clusterwise regression; Noise cluster; Robustness

1. Introduction

Fuzzy regression analysis is a fuzzy (or possibility) type of classical regression analysis. It is used in evaluating the functional relationship between the dependent and independent variables in a fuzzy environment. Tanaka et al. [11] initiated a study in fuzzy linear regression analysis which considered the parameter estimation of models as linear programming problems. Subsequently, there have been many other studies which can be roughly divided into two approaches: linear-programming-based methods [8–11] and fuzzy least-squares methods [1,5,14]. Most of these fuzzy regression models are considered with fuzzy outputs and fuzzy parameters but non-fuzzy (crisp) inputs. In this paper fuzzy linear regression models with fuzzy outputs, fuzzy parameters, and fuzzy inputs are considered.

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

0165-0114/02/$ - see front matter
PII: S0165-0114(01)00066-5

Sakawa and Yano [8] proposed a fuzzy parameter estimation for the fuzzy linear regression (FLR) model

$$Y_j = A_0 + A_1 X_{j1} + \cdots + A_k X_{jk}, \quad j = 1, \ldots, n,$$

where both input data $X_{j1}, \ldots, X_{jk}$ and output data Y_j are fuzzy, by using three indices for the equalities between two fuzzy numbers M and N as

$$Pos(M = N) = \sup_x \min\{\mu_M(x), \mu_N(x)\},$$

$$Nes(M \subset N) = \inf_x \max\{1 - \mu_M(x), \mu_N(x)\},$$

$$Nes(M \supset N) = \inf_x \max\{\mu_M(x), 1 - \mu_N(x)\},$$

where $\mu_M(x)$ and $\mu_N(x)$ are membership functions of M and N, and Pos and Nes are short for Possibility and Necessity. Then three types of multiobjective programming problems were formulated for the parameter estimation of FLR models along with a linear-programming-based approach. This multicriterial analysis of FLR models provided a good method of parameter estimation by using the vagueness of the model via some indices of inclusion relations. Alternatively, a fuzzy least-squares approach directly uses information included in the input–output data set and considers the measure of best fitting based on distance under fuzzy consideration. Fuzzy least-squares are fuzzy extensions of ordinary least-squares. In this paper two types of fuzzy least-squares are proposed as the parameter estimation for the FLR model $Y_j = A_0 + A_1 X_{j1} + \cdots + A_k X_{jk}$, $j = 1, \ldots, n$. These proposed methods have proved to be a good alternative approach to a multiobjective programming approach.

Section 2 presents the first method, called approximate-distance fuzzy least-squares. In order to treat heterogeneous data sets and detect outliers, a clusterwise fuzzy least-square with a noise cluster is also presented. In Section 3, another fuzzy least-squares method, called an interval-distance fuzzy least-square, is proposed. Numerical results are presented in Section 4. Finally, conclusions are made in Section 5.

2. Approximate-distance fuzzy least-squares

The following FLR model is considered:

$$Y_j = A_0 + A_1 X_{j1} + \cdots + A_k X_{jk}, \quad j = 1, \ldots, n, \tag{2.1}$$

where outputs $Y_j = (m_{y_j}, \alpha_{y_j}, \beta_{y_j})_{LR}$, inputs $X_{ji} = (m_{x_{ji}}, \alpha_{x_{ji}}, \beta_{x_{ji}})_{LR}$ and parameters $A_i = (m_{a_i}, \alpha_{a_i}, \beta_{a_i})_{LR}$, $i = 1, \ldots, k$, $j = 1, \ldots, n$ so that the notion $M = (m, \alpha, \beta)_{LR}$ is an LR-type fuzzy number with its membership function

$$\mu_M(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right) & \text{for } x \leqslant m \\ R\left(\dfrac{x-m}{\beta}\right) & \text{for } x \geqslant m \end{cases} \qquad (\alpha > 0,\ \beta > 0).$$

In $M = (m, \alpha, \beta)_{LR}$, m is called the mean value, and α and β are called the left and right spreads, respectively. If $L(x) = R(x) = 1 - x$, then $M = (m, \alpha, \beta)_{LR}$ is called a triangular fuzzy number, denoted by $M = (m, \alpha, \beta)_T$.

The difficulty in treating model (2.1) of fuzzy input–output data is that $A_i X_{ji}$ may not be of LR-type. Although the product of two LR-type fuzzy numbers may not be an LR-type fuzzy number, Dubois and Prade [6] presented an approximation form. Therefore, Yang and Ko [14] suggested an approximation type of fuzzy least-squares for an extension to their results. In this section, the idea of approximation is used to present an algorithm for parameter estimation of the FLR model (2.1).