Expert Systems with Applications 38 (2011) 13632–13640

Contents lists available at ScienceDirect

Expert Systems with Applications

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

The construction of fuzzy least squares estimators in fuzzy linear regression models

Hsien-Chung Wu

Department of Mathematics, National Kaohsiung Normal University, Kaohsiung 802, Taiwan

ARTICLE INFO

Keywords:
Confidence interval
Fuzzy numbers
Least squares estimator
Optimization
Regression
Testing hypothesis

ABSTRACT

A new concept and method of imposing imprecise (fuzzy) input and output data upon the conventional linear regression model is proposed. Under the considerations of fuzzy parameters and fuzzy arithmetic operations (fuzzy addition and multiplication), we propose a fuzzy linear regression model which has the similar form as that of conventional one. We conduct the h-level (conventional) linear regression models of fuzzy linear regression model for the sake of invoking the statistical techniques in (conventional) linear regression analysis for real-valued data. In order to determine the sign (nonnegativity or nonpositivity) of fuzzy parameters, we perform the statistical testing hypotheses and evaluate the confidence intervals. Using the least squares estimators obtained from the h-level linear regression models, we can construct the membership functions of fuzzy least squares estimators via the form of "Resolution Identity" which is well-known in fuzzy sets theory. In order to obtain the membership degree of any given estimate taken from the fuzzy least squares estimator, optimization problems have to be solved. We also provide two computational procedures to deal with those optimization problems.

1. Introduction

In the real world, the data sometimes cannot be recorded or collected precisely due to human errors, machine errors or some other unexpected situations. For instance, the water level of a river cannot be measured in an exact way because of the fluctuation, and the temperature in a room is also not able to be measured precisely because of similar reasons. With this situation, fuzzy sets theory is naturally an appropriate tool in statistical models when fuzzy data have been observed. The more appropriate way to describe the water level is to say that the water level is around 30 m. The phrase "around 30 m" can be regarded as a fuzzy number $\widetilde{30}$. This is the main concern of this paper.

Since Zadeh (1965) introduced the concept of fuzzy sets, the applications of considering fuzzy data to the regression models have been proposed in the literature. Tanaka, Uejima, and Asai (1982) initiated this research topic. They also generalized their approaches to more general models in Tanaka and Warada (1988, 1989, 1991). The book on fuzzy regression analysis edited by Kacprzyk and Fedrizzi (1992) gave an insightful survey. Chang and Ayyub (2001) gave the differences between the fuzzy regression and conventional regression analysis and Kim, Moskowitz, and Koksalan (1996) also compared both fuzzy regression and statistical regression conceptually and empirically.

In the approach of Tanaka et al. (1982), they considered the L–R fuzzy data and minimized the index of fuzziness of the fuzzy linear regression model. Redden and Woodall (1994) compared various fuzzy regression models and gave the difference between the approaches of fuzzy regression analysis and conventional regression analysis. They also pointed out some weakness of the approaches proposed by Tanaka et al. Chang and Lee (1994) also pointed out another weakness of the approaches proposed by Tanaka et al. Peters (1994) introduced a new fuzzy linear regression models based on Tanaka's approach by considering the fuzzy linear programming problem. Moskowitz and Kim (1993) proposed a method to assess the H-value in a fuzzy linear regression model proposed by Tanaka et al. Wang and Tsaur (2000) also proposed a new model to improve the predictability of Tanaka's model. In this paper, we propose a fuzzy linear regression model, and then the h-level linear regression models will be created by taking the h-level set of fuzzy linear regression model. We shall see that the h-level linear regression models are conventional linear regression models. Therefore, the statistical techniques proposed in the conventional linear regression analysis can be invoked to discuss the h-level linear regression models.

For the least squares sense, Chang (2001) proposed a method for hybrid fuzzy least squares regression by defining the weighted fuzzy-arithmetic and using the well-accepted least squares fitting criterion. Celminš (1987, 1991) proposed a methodology for the fitting of differentiable fuzzy model function by minimizing a least squares objective function. Chang and Lee (1996) proposed a fuzzy regression technique based on the least squares approach to estimate the modal value and the spreads of L–R fuzzy number. Jajuga (1986) calculated the linear fuzzy regression coefficients using a generalized version of the least squares method by considering the fuzzy classification of a set of observations and obtaining the

E-mail addresses: hcwu@nknucc.nknu.edu.tw, hcwu@im.ncnu.edu.tw

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.04.131

homogeneous classes of observations. In this paper, the least squares estimators will be obtained from the h-level linear regression models. Using these least squares estimators, we can construct a fuzzy least squares estimators via the form of "Resolution Identity" which is introduced by Zadeh et al. (1975) and is well-known in fuzzy sets theory.

For optimization approach, Sakawa and Yano (1992) introduced three indices for equalities between fuzzy numbers. From these three indices, three types of multiobjective programming problems were formulated. Tanaka and Lee (1998) used the quadratic programming approach to obtain the possibility and necessity regression models simultaneously. The advantage of adopting a quadratic programming approach is to be able to integrate both the property of central tendency in least squares and the possibilistic property in fuzzy regression. In this paper, in order to obtain the membership value (confidence degree) of any given estimate taken from the fuzzy least squares estimator, the optimization problems have to be solved. We also provide two computational procedures to solve those optimization problems.

There are also some other interesting articles concerning the fuzzy regression analysis. Näther (1997, 2000), Näther and Albrecht (1990) and Körner et al. (1998) introduced the concept of random fuzzy sets (fuzzy random variables) into the linear regression model, and developed an estimation theory for the parameters. Dunyak and Wunsch (2000) described a method for nonlinear fuzzy regression using a special training technique for fuzzy number neutral networks. Kim and Bishu (1998) used a criterion of minimizing the difference of the membership degrees between the observed and estimated fuzzy numbers. Yager (1982) used a linguistic variable to represent imprecise information for the regression models. Bárdossy (1990) proposed many different measures of fuzziness which must be minimized with respect to some suggested constraints.

In Section 2, we give some properties of fuzzy numbers. In Section 3, The techniques for solving fuzzy linear regression problems are proposed. We shall focus on the h-level linear regression models of fuzzy linear regression model, and then apply the conventional linear regression techniques to solve the h-level linear regression models. The membership functions of fuzzy least squares estimators in fuzzy linear regression model will be constructed according to the form of "Resolution Identity" in fuzzy sets theory. In Section 4, we shall develop two computational procedures to obtain the membership degree of any given estimate taken from the fuzzy least squares estimators. We also provide an example to clarify the theoretical results, and show the possible applications in linear regression analysis for imprecise data.

2. Fuzzy numbers

Let X be a universal set and A be a subset of X. Then the indicator (characteristic) function 1_A defined by

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

can be used to represent the subset A of X. A fuzzy subset $\widetilde{A}$ of X proposed by Zadeh (1965) is defined by its membership function $\xi_{\widetilde{A}} : X \to [0,1]$. We see that the concept of membership function is an extension of the indicator function 1_A of A, since the indicator function 1_A can also be regarded as a membership function of $\widetilde{A}$. In this case, the indicator function is sometimes called as a crisp function in the fuzzy sets theory.

Let X be endowed with a topological structure τ. We denote by $\widetilde{A}_h = \{x : \xi_{\widetilde{A}}(x) \geqslant h\}$ the h-level set of $\widetilde{A}$ for $h \in (0,1]$, and $\widetilde{A}_0$, called as support of $\widetilde{A}$, is taken as the closure of the set $\{x : \xi_{\widetilde{A}}(x) > 0\}$. $\widetilde{A}$ is called a normal fuzzy set if there exists an x such that $\xi_{\widetilde{A}}(x) = 1$, and is called a convex fuzzy set if $\xi_{\widetilde{A}}(\lambda x + (1-\lambda)y) \geqslant \min\{\xi_{\widetilde{A}}(x), \xi_{\widetilde{A}}(y)\}$ for $\lambda \in [0,1]$, i.e., $\xi_{\widetilde{A}}$ is a quasi-concave function (Ref. Royden (1968) & Rudin (1986)).

Throughout this paper, the universal set X is assumed to be a real number system $\mathbb{R}$ endowed with a usual topology. Let f be a real-valued function defined on $\mathbb{R}$. Then f is said to be upper semicontinuous if $\{x: f(x) \geqslant h\}$ is a closed set for each h. Or equivalently, f is upper semicontinuous at y if and only if $\forall \epsilon > 0, \exists \delta > 0$ such that $|x - y| < \delta$ implies $f(x) < f(y) + \epsilon$ (Ref. Royden (1968) & Rudin (1986)).

Let $\tilde{a}$ be a fuzzy subset of $\mathbb{R}$. Then $\tilde{a}$ is called a fuzzy number if the following conditions are satisfied:

- $\tilde{a}$ is a normal and convex fuzzy subset of $\mathbb{R}$.
- Its membership function $\xi_{\tilde{a}}$ is upper semicontinuous.
- The h-level set $\tilde{a}_h$ is bounded for each $h \in [0,1]$.

From Zadeh (1965), $\widetilde{A}$ is a convex fuzzy subset of $\mathbb{R}$ if and only if its h-level set $\widetilde{A}_h = \{x : \xi_{\widetilde{A}}(x) \geqslant h\}$ is a convex subset of $\mathbb{R}$ for all $h \in [0,1]$ (A is a convex subset of $\mathbb{R}$ if and only if $tx_1 + (1-t)x_2 \in A$ for any $x_1, x_2 \in A$ and $t \in (0,1)$). Therefore if $\tilde{a}$ is a fuzzy number, then the h-level set $\tilde{a}_h$ is a closed and bounded subset of $\mathbb{R}$ from conditions (ii) and (iii), and a convex subset of $\mathbb{R}$ from conditions (i); that is, $\tilde{a}_h$ is a closed interval in $\mathbb{R}$. The h-level set of $\tilde{a}$ is then denoted by $\tilde{a}_h = [\tilde{a}_h^L, \tilde{a}_h^U]$. We also see that $\tilde{a}_h^L$ and $\tilde{a}_h^U$ are continuous with respect to h (i.e., $l(h) \equiv \tilde{a}_h^L$ and $u(h) \equiv \tilde{a}_h^U$ are continuous functions for a fixed fuzzy number $\tilde{a}$), since its membership function is upper semicontinuous. The following proposition is useful for further discussions.

Proposition 2.1 (Zadeh et al. (1975)(Resolution Identity)). *Let $\widetilde{A}$ be a fuzzy subset of X with membership function $\xi_{\widetilde{A}}$, and the h-level set $\widetilde{A}_h = \{x : \xi_{\widetilde{A}}(x) \geqslant h\}$ of $\widetilde{A}$ be given. Then the membership function can be represented as*

$$\xi_{\widetilde{A}}(x) = \sup_{h \in [0,1]} h \cdot 1_{\widetilde{A}_h}(x),$$

where $1_{\widetilde{A}_h}$ is the indicator function of (crisp) set $\widetilde{A}_h$.

$\tilde{a}$ is called a crisp number with value m if its membership function is given by

$$\xi_{\tilde{a}}(x) = \begin{cases} 1 & \text{if } x = m \\ 0 & \text{otherwise.} \end{cases}$$

It is denoted by $\tilde{a} \equiv \tilde{1}_{\{m\}}$. We also see that $(\tilde{1}_{\{m\}})_h^L = (\tilde{1}_{\{m\}})_h^U = m$ for all $h \in [0,1]$.

Let $\tilde{a}$ be a fuzzy number. Then $\tilde{a}$ is called nonnegative if $\xi_{\tilde{a}}(x) = 0$ for all $x < 0$ and nonpositive if $\xi_{\tilde{a}}(x) = 0$ for all $x > 0$. It is obvious that $\tilde{a}_h^L$ and $\tilde{a}_h^U$ are nonnegative real numbers for all $h \in [0,1]$ if $\tilde{a}$ is a nonnegative fuzzy number, and $\tilde{a}_h^L$ and $\tilde{a}_h^U$ are nonpositive real numbers for all $h \in [0,1]$ if $\tilde{a}$ is a nonpositive fuzzy number.

Let "$\odot$" be any binary operation $\oplus$ or $\otimes$ between two fuzzy numbers $\tilde{a}$ and $\tilde{b}$. The membership function of $\tilde{a} \odot \tilde{b}$ is defined by

$$\xi_{\tilde{a}\odot\tilde{b}}(z) = \sup_{x \circ y = z} \min\left\{\xi_{\tilde{a}}(x), \xi_{\tilde{b}}(y)\right\}$$

using the extension principle in Zadeh et al. (1975), where the operations $\odot = \oplus$ or $\otimes$ corresponding to the operations $\circ = +$ or $\times$. Then we have the following well-known results.

Proposition 2.2. *Let $\tilde{a}$ and $\tilde{b}$ be two fuzzy numbers. Then $\tilde{a} \oplus \tilde{b}$ and $\tilde{a} \otimes \tilde{b}$ are also fuzzy numbers. Furthermore, we have*

$$(\tilde{a} \oplus \tilde{b})_h = \left[\tilde{a}_h^L + \tilde{b}_h^L, \tilde{a}_h^U + \tilde{b}_h^U\right]$$

and

$$(\tilde{a} \otimes \tilde{b})_h = \left[\min\left\{\tilde{a}_h^L\tilde{b}_h^L, \tilde{a}_h^L\tilde{b}_h^U, \tilde{a}_h^U\tilde{b}_h^L, \tilde{a}_h^U\tilde{b}_h^U\right\}, \max\left\{\tilde{a}_h^L\tilde{b}_h^L, \tilde{a}_h^L\tilde{b}_h^U, \tilde{a}_h^U\tilde{b}_h^L, \tilde{a}_h^U\tilde{b}_h^U\right\}\right].$$